Rogers J, Rector A (1997). *Terminological Systems: Bridging the Generation Gap.* Annual Fall Symposium of American Medical Informatics Association. Nashville TN, Hanley & Belfus Inc. Philadelphia PA: 610-614.

Exhibit B

Terminological Systems: Bridging the Generation Gap

JE Rogers AL Rector
Medical Informatics Group, Department of Computer Science, University of Manchester, UK
http://www.cs.man.ac.uk/mig/giu/ or e-mail galen@cs.man.ac.uk

A rigorous formal description of the intended behaviour of a compositional terminology, implemented as a software engine, enables advanced, powerful semantic processing techniques to assist in the building of a large terminology. Use of an intermediate representation derived from such a formalism enables authors to work in an apparently less formal environment, accessing these techniques at one remove.

INTRODUCTION

Developers of terminologies specifically designed for medical computer applications are increasingly exploring alternatives to the enumerative techniques embodied by traditional schemes such as ICD[1] or READ version 1 or 2. Expressivity of such schemes is limited by whether appropriate, specific terms already exist. Existing terminologies such as SNOMED[2], and many currently in development (e.g. DICOM SNOMED Microglossary, LOINC, ICNP[3], READ 3.1[4]), have adopted compositional techniques: increased expressivity is achieved by fashioning descriptions from structured collections of more basic terms.

Compositionality increases flexibility: a common clinical requirement is for sets of highly detailed terms in a particular specialised medical sub-domain - perhaps for research or audit purposes. Users of enumerative schemes must either wait for them to be included in the next major central revision or (more commonly) make *ad hoc* local additions. A compositional scheme enables principled local extension, by making new compositions. The need for genuinely new atomic terms is, therefore, much reduced.

European standardisation work reflects this move to compositional techniques. The European Committee for Standardisation (CEN) has produced several standards and pre-standards following ENV 12264[5], itself a pre-standard for representing terminologies as a semantic network.

Existing enumerative schemes are termed 'first generation' terminology systems by Rossi Mori[6]. In his study of compositional schemes in development he identifies four common components: a *categorial structure*, a *cross-thesaurus*, a *family of lists* and a *knowledge base of dissections*. Systems where all four components are well developed - Rossi Mori's 'second generation' - acquire new capabilities of semantic processing. These include dynamic re-organisation of compositions, support for structured data entry, the ability to automatically generate extensions and dynamic cross-referencing between other schemes.

However, Rossi Mori notes that developing the four components and the resulting scheme must be an iterative process. Further, development of one component often complements, but may also depend upon, development of another. These dependencies may initially be expressed as a set of manually applied rules and checks. However, as the system and its dependencies become progressively more complex, it ceases to be possible to maintain integrity or coherence through human processing power alone.

Further progress requires formal encapsulation of the system's intended behaviour in a software engine. Systems including such an engine - Rossi Mori's 'third generation' systems - constrain and guide all user interaction according to this formalism. Further enhancements of semantic processing power are gained, but knowledge authoring becomes more demanding: the scheme, its terms and the formalism become so interdependent as to be inseparable and the whole becomes essentially a piece of software.

GALEN-IN-USE

GALEN-IN-USE is a European Union funded project to develop tools and methods to assist in the collaborative construction and maintenance of compositional surgical procedure classifications. This paper describes how results from the previous GALEN project - the GRAIL formalism[7], GALEN Common Reference Model (CRM)[8,9,10], High Level Ontology[11] and Terminology Servers[12] - are providing 'third generation' system support for this task.

Taking part in the initial phase are four national coding and classification centres: WCC (Netherlands), SPRI (Sweden), CNR (Italy) and University of Ste. Etienne (France). During the project, conceptual representations of some 15,000 individual surgical procedures will be produced using the GRAIL formalism and integrated into the existing GALEN Common Reference Model[7,9,10,11].

GALEN and CEN ENV 1828

The relationship between GALEN and 'second generation' systems is illustrated by the GALEN approach to CEN ENV 1828[13], a pre-standard proposing a compositional structure for classifications of surgical procedures. The CEN schema (figure 1) reflects the way the terms are used in language. Our experience has been that a conceptual model has slightly different requirements.[14] GALEN's schema must both support automatic classification and also integrate with an existing model which permits indefinite nesting of anatomical sublocation. These different treatments are

Figure 1: CEN ENV 1828 schema for surgical procedures

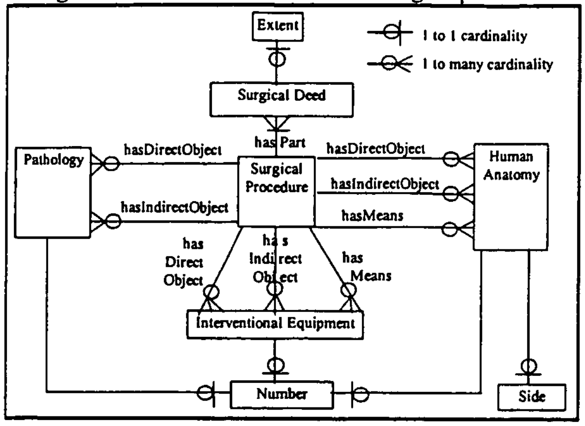
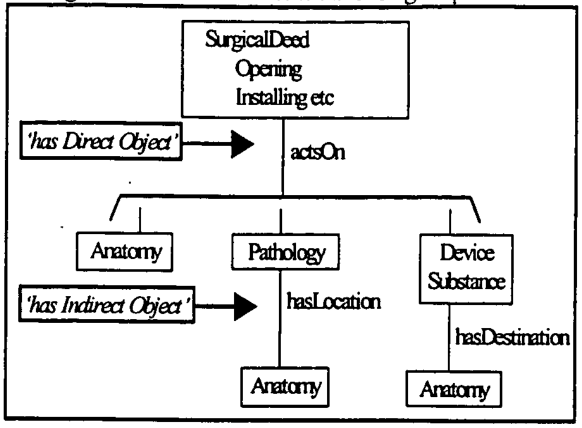


Figure 2: Basic GALEN schema for surgical procedures



illustrated by the GALEN interpretation of the section in the normative part of ENV 1828 which states:

A surgical procedure must have anatomy either as a direct or an indirect object

Within the GALEN Common Reference Model (CRM), neither the indirect nor the direct object is linked directly to the procedure. Instead, the direct object is always linked to the surgical deed itself:

(Removing which actsOn Kidney) name Nephrectomy.

A more significant difference in treatments concerns the indirect object. In the CEN schema, the notion of 'excision of a kidney cyst' would be expressed as:

(SurgicalProcess:)*
– (hasPart) – (SurgicalDeed: Removal)
– (hasDirectObject) – (Pathology: Cyst)
– (hasIndirectObject) – (Anatomy: Kidney)

However, in the Common Reference Model we are able to specialise [Cyst] according to its location:

(Cyst which hasLocation Kidney) name KidneyCyst.

If the CEN schema were followed, the constraining mechanisms in GRAIL could not prevent construction of obviously nonsense compositions such as 'removal of a renal cyst from the thyroid':

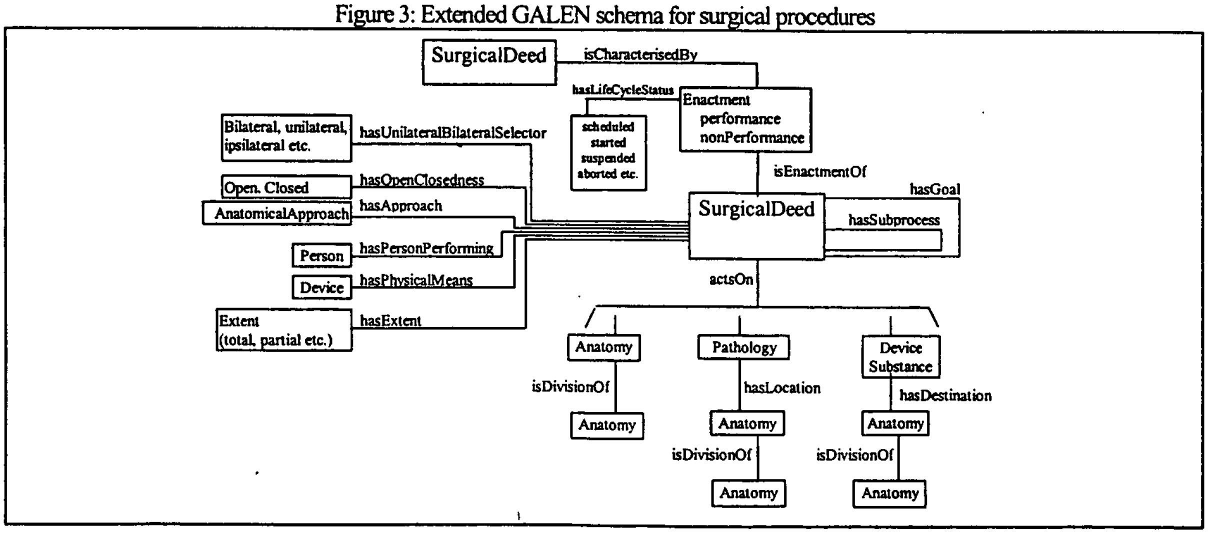
(Removing)
– (actsOn) – [(Cyst) – (hasLocation) – (Kidney)]
– (hasIndirectObject) – (Thyroid)

In the CRM, therefore, the indirect object is attached indirectly to the deed, via the direct object, thus:

(Removing which actsOn (Cyst which hasLocation Kidney)).

These changes result in a basic GALEN schema for surgical procedures (figure 2). This has subsequently been expanded to increase expressivity and to integrate it with other modelling schemata already present in the Common Reference Model (figure 3).

Figure 3: Extended GALEN schema for surgical procedures



Rogers J, Rector A (1997). *Terminological Systems: Bridging the Generation Gap.* Annual Fall Symposium of American Medical Informatics Association. Nashville TN, Hanley & Belfus Inc. Philadelphia PA: 610-614.

AN INTERMEDIATE REPRESENTATION

GRAIL, the GALEN representation language, is necessarily complex - as would be any other 'third generation' representation. For the GALEN-IN-USE project, more than 20 clinicians were recruited across four countries to perform the analysis of original surgical procedure code rubrics into conceptual representations. However, few had any prior experience of GRAIL or the Common Reference Model.

To circumvent this problem we devised an intermediate representation[15] nearer to a 'second generation' system. It is structurally simpler than GRAIL, but may subsequently be automatically expanded into GRAIL. This expansion is possible because the design of the intermediate representation deliberately echoes that of the Common Reference Model. For example, the schema for surgical procedures in the intermediate representation is a systematic simplification of the corresponding extended GALEN schema. This enables automatic 'de-simplification' to occur when the dissections are expanded into GRAIL. Rogers has described this expansion process[16] and the GALEN software tools (TIGGER and SPET) which support it.

The intermediate representation is broadly similar to those used by the CANON group or the MEDS.[17,18,19,20] It is characterised by:

- a grammar defining a layout, or 'template', for well-formed representations.
- a relatively small set of semantic links (ACTS_ON, IS_PART_OF), compared to the GALEN CRM;
- a domain ontology specific to the surgical domain. The atomic terms (leg, excising, tumour etc.) are known as 'descriptors' and are explicitly typed by one of a small number of descriptor classes (e.g. anatomy, deed, lesion);
- a small set of constraints to control which links may be used with which descriptor classes.

Domain experts in the centres work from existing local coding schemes (WCC, NCSP etc.) to scope their task. Rubrics from these schemes are manually analysed to give, initially, a natural language paraphrase of what the expert believes the rubric means. A conceptual representation of each such paraphrase is then produced using the intermediate representation. The result of this two-step analysis is called a 'dissection' of the rubric. Each dissection has a header section which contains information about the original rubric and coding scheme. This is followed by the conceptual representation itself, introduced by the MAIN keyword. Semantic links are capitalised, descriptors are in lower case. Below is an example of a completed dissection:

```
RUBRIC "Insertion of intercostal catheter for drainage"
PARAPHRASE "Insertion of intercostal catheter in pleural
space for drainage"
SOURCE "ICD-9-CM" CODE "34.04"
MAIN inserting
    ACTS_ON catheter
    HAS_APPROACH intercostal route
    HAS_DESTINATION pleural space
MOTIVATED_OVERALL_BY draining
    ACTS_ON substance
        HAS_LOCATION pleural space
```

A GRAIL expansion from this dissection is automatically generated (below). The expansion algorithm requires that the primitive descriptors and links in the intermediate representation are given context dependent mappings to primitive or composed concepts and attributes in the Common Reference Model, as described by Rogers.[16]

```
[(SurgicalDeed whichG <
   isMainlyCharacterisedBy
     (performance whichG isEnactmentOf
       (Inserting which <
           hasSpecificSubprocess
(SurgicalApproaching whichG hasPhysicalMeans
  (Route which passesThrough IntercostalSpace))
           isActedOnSpecificallyBy
(Transport whichG hasSpecificConsequence
  (Displacement whichG hasBetaConnection PleuralCavity))
           playsClinicalRole SurgicalRole
           actsSpecificallyOn Catheter>))
  hasSpecificGoal (Draining which <
    playsClinicalRole SurgicalRole
    actsOn (Substance whichG hasLocation PleuralCavity)>) ]
  hasProjection
(('ICD-9-CM' schemeVersion 'default') code '34.04' 'code');
  extrinsically hasDissectionRubric
'ICD-9-CM 34.04 Insertion of intercostal catheter for drainage'.
```

ADDED VALUE OF GALEN

The GALEN intermediate representation is similar to a 'second generation' system. However, it results from a systematic simplification of a 'third generation' system rather than a gradual increase in sophistication of a 'first generation' enumerative system. This approach facilitates our knowledge authoring process whilst still allowing 'third generation' techniques to be exploited to build, maintain and validate the corpus and, ultimately, deliver it to end users. Four techniques, not applicable to 'second generation' systems or the intermediate representation directly, are fundamental to our authoring process:

- Automated semantic normalisation and canonisation
- Automated and dynamic classification of compositions
- Automated maintenance of fixed knowledge database
- Automated generation of natural languages

Semantic Normalisation

Different authors, analysing the same rubrics, produce different dissections. These differences divide into those which are semantically equivalent, those semantically divergent and those which represent semantic error. The expansion of dissections into GRAIL provides several different stages at which normalisation can occur. For example, differences of semantic equivalence such as varying encapsulation may be automatically normalised. A separate mechanism rejects many semantic errors:

Normalising varying encapsulation: in the rubric 'excision of lobe of lung', one author may determine that {lobe of lung} is an appropriate primitive descriptor, whilst another may choose the decomposition {lobe IS_PART_OF lung}. The expansion into GRAIL normalises both into:

(Lobe which isSolidDivisionOf Lung).

because of the following previously declared mappings:

Descriptor / Link	GRAIL Mapping
lobe of lung	Lobe which isSolidDivisionOf Lung
lobe	Lobe
lung	Lung
IS_PART_OF	isSolidDivisionOf

Rejecting semantic error: The intermediate representation includes only a limited set of constraints controlling which classes of descriptor may be combined with which links. A richer set of constraints exists in the CRM, and these are brought to bear when a dissection is expanded into GRAIL. Thus {fracturing ACTS_ON temperature} is permitted in the intermediate representation, but rejected at the GRAIL expansion stage.

Semantic divergence: Differences of opinion between experts regarding what rubrics actually mean must remain problems for the experts to resolve. However, the other techniques discussed here combine to assist the domain experts in identifying when they do not agree.

Automatic classification

GRAIL expansions of the dissections are automatically classified according to the principles of the GRAIL formalism. Knowledge already present in the CRM affects this classification; for example, 'Operation on the Heart' subsumes 'Repair of Mitral Valve' because the anatomy model already knows the mitral valve is part of the heart.

Where a dissection has *not* been classified as expected, the task is to identify why. With the 'noise' of semantic equivalence removed through normalisation, the remaining causes are semantic divergence, and omissions or errors in the pre-existing knowledge base. Automated analysis, according to the formalism, of the relationships between expansions of dissections can answer questions such as 'why is this classified here?' and 'what should I change to have it classified there?'.

Automatic classification further ensures that the twin hierarchies of composed deeds and of the objects they act on must inevitably be exactly parallel, since one is derived formally from the other. Maintaining this 'parallelism' is presently commonly undertaken manually in other 'second generation' systems, (e.g. the READ 3.1 Thesaurus[4]).

Maintenance of the knowledge database

To hold a fixed form of the knowledge base, local implementations of compositional systems may need to instantiate 'artefact' concepts as well as the compositions originally provided by authors. This might be necessary to fit the knowledge base within a particular persistent data structure, (as occurs in the READ 3.1 Thesaurus[4]) or to optimise a classification or search algorithm.

In a GALEN system, knowledge authoring is decoupled from any particular implementation of the knowledge base. The local implementation determines for itself what it needs to instantiate, and is able to export the knowledge base to other implementations where the requirements for instantiated concepts may be different.

Machine language generation

Early experiments provided the dissection authors with a display of their original scheme rubrics, ordered into a hierarchy according to the automatic classification of the GRAIL expansions. However, the original rubric is not always a satisfactory proxy for the dissection itself. The semantic information which directly determines the classification is hidden, and identifying the cause of an inappropriate classification from this presentation alone is not possible. Similarly, browsing the hierarchy of the GRAIL concepts themselves displays too much information, in too abstract a form, to be directly useful.

GALEN tools can generate from a GRAIL composition a natural language string which reflects the semantics of that composition.[21] Browsing hierarchies of these strings, in an editing environment which links them directly to their originating rubrics, dissections or GRAIL expansions, is expected to form a powerful QA tool.

RESULTS AND FUTURE DEVELOPMENTS

More than 3000 original rubrics, in the fields of orthopaedics, urology, cardiology and gastroenterology have so far been dissected using the intermediate representation. These have subsequently been expanded into GRAIL and classified within the Common Reference Model. Generation of Natural Language phrases for the results is now possible in four European languages, though the lexicons are not yet complete. Future experiments will examine delivering the corpus to the participating centres as either a first, second or third generation system according to local requirements.

CONCLUSION

'Third generation' systems, such as GALEN, offer advanced semantic processing techniques. We have shown the added value of using these to help build large and coherent terminologies. However, authoring compositional representations directly in a formalism such as GRAIL is time consuming and requires special skills.

An intermediate representation can bridge between the generations: 'third generation' system advantages can be gained whilst authoring effort remains closer to that required for 'second generation' systems. Existing standards can be extended or adapted to support this activity. A prerequisite is an automatic transformation between this representation and the formalism, and between the formalism and natural language.

Acknowledgements

With thanks to all in the GALEN-IN-USE consortium.
GALEN-IN-USE is funded as part Framework IV of the EC Healthcare Telematics research program.

References

1. World Health Organisation. International Classification of Diseases, 9th Revision. Geneva: WHO, 1977
2. Cote RA, Rothwell DJ, Palotay JL, Beckett RS, Brochu L (eds), The Systematised Nomenclature of Human and Veterinary Medicine: SNOMED International, College of American Pathologists , Northfield, IL:, 1993, 3rd edition
3. Mortensen RA (ed) The International Classification for Nursing Practice ICNP with TELENURSE introduction. Copenhagen: The Danish Institute for Health and Nursing Research, 1996
4. Price C, et al. Anatomical Characterisation of Surgical Procedures in the Read Thesaurus. JAMIA 1996; symp. Suppl.;110-114
5. CEN ENV 12264:1995. Medical Informatics - Categorial structure of systems of concepts - Model for representation of semantics. Brussels: CEN, 1995
6. Rossi Mori A, Consorti F, Galeazzi E, (1997) Standards to support development of terminological systems for healthcare telematics. (proceedings of IMIA Working Group 6 meeting, Jacksonville, Florida)
7. Rector A, and Nowlan WA (1993). The GALEN Representation and Integration Language (GRAIL) Kernel, Version 1. The GALEN Consortium for the EC AIM Programme. (Available from Medical Informatics Group, University of Manchester).
8. Rector A (1994). Compositional models of medical concepts: towards re-usable application-independent medical terminologies. Knowledge and Decisions in Health Telematics P. Barahona and J. Christensen (ed). IOS Press. 133-142.
9. Rector A, Gangemi A, Galeazzi E, Glowinski A and Rossi-Mori A (1994). The GALEN CORE Model Schemata for Anatomy: Towards a re-usable application-independent model of medical concepts. Twelfth International Congress of the European Federation for Medical Informatics, MIE-94, Lisbon, Portugal, 229-233
10. Rector A (1995). Coordinating taxonomies: Key to re-usable concept representations. Fifth conference on Artificial Intelligence in Medicine Europe (AIME '95), Pavia, Italy, Springer. 17-28.
11. Rector A, Rogers JE ,Pole P (1996) The GALEN High Level Ontology. Fourteenth International Congress of the European Federation for Medical Informatics, MIE-96, Copenhagen, Denmark
12. Rector A, Solomon WD, Nowlan WA and Rush T (1995). A Terminology Server for Medical Language and Medical Information Systems. Methods of Information in Medicine, Vol. 34, 147-157
13. CEN ENV 1828:1995 Health care informatics - Structure for classification and coding of surgical procedures. Brussels: CEN, 1995
14. Ceuster W, Beukens F, De Moor G, Waagmeester A (1997). The Distinction between Linguistic and Conceptual Semantics in Medical terminology and its Implications for NLP-Based Knowledge Acquisition. (proceedings of IMIA Working Group 6 meeting, Jacksonville, Florida)
15. Gaines BR, Shaw ML and Woodward JB (1993). Modelling as Framework for Knowledge Acquisition Methodologies and Tools. International Journal of Intelligent Systems 8(2): 155-168.
16. Rogers JE, Solomon WD et al. (1997) Rubrics to Dissections to GRAIL to Classifications. Fifteenth International Congress of the European Federation for Medical Informatics, MIE-97 Thessaloniki, Greece
17. Campbell KE, Das AK and Musen MA (1994). A logical foundation for representation of clinical data. JAMIA 1(3): 218-232.
18. Cimino J (1994). Controlled Medical Vocabulary Construction: Methods from the Canon Group. Journal of the American Medical Informatics Association 1(3): 296-197.
19. Evans D (1988). Pragmatically-structured, lexical-semantic knowledge bases for unified medical language systems. Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care, Washington DC, IEEE Computer Society Press: 169-173.
20. Huff S and Warner H (1990). A comparison of Meta-1 and HELP terms: implications for clinical data. Fourteenth Annual Symposium on Computer Applications in Medical Care (SCAMC-90), Washington DC, iEEE Computer Society Press: 166-169.
21. Baud RH, Rassinoux A-M, Lovis C, Wagner J et al. Knowledge Sources for Natural Language Processing. In: Cimino JJ (ed) Proceedings of the 1996 AMIA Annual Fall Symposium Philadelphia: Hanley & Belfus, Inc. 1996: 70-84

Exhibit C

Compositional Concept Representation Using SNOMED: Towards Further Convergence of Clinical Terminologies

Kent A. Spackman, M.D., Ph.D.[1,2], Keith E. Campbell, M.D., Ph.D.[3],
[1]College of American Pathologists, Northfield, IL;
[2]Oregon Health Sciences University, Portland, OR;
[3]Kaiser-Permanente, Oakland, CA

This paper describes several approaches to the expression and coding of clinical concepts as composites of elementary entities, and describes an approach based on SNOMED RT that may permit further convergence of clinical terminology efforts. We explain the shortcomings of previous approaches to compositional concept representation, as well as the reasons for SNOMED's current approach, which adopts a foundation based in description logics (DLs). The DL model has many advantages: it establishes a formal semantics for SNOMED assertions and suggests a syntax; it provides a basis for understanding expressiveness and computational complexity, through correspondence with known results from DLs; and it helps to clarify the relationships among existing concept representation methods in SNOMED, NHS Clinical Terms (formerly the Read Codes), and GALEN, making a path to convergence more clear.

INTRODUCTION

One of the main reasons to create a comprehensive clinical terminology is to facilitate the accurate representation of clinical detail, allowing accurate storage, retrieval and analysis of patient data. Natural descriptions of clinical details are richly varied, and it would be practically impossible to enumerate them all. It has long been recognized that it would be desirable to have a model that allows composition of clinical concepts from atomic elements [1,12]. SNOMED [2] has always allowed compositional encoding, but many authors have recognized shortcomings in SNOMED's compositional concept representation. For example, some have called for a syntax, and others have noted that it is possible to represent the same concept with more than one unique combination of codes. [3] We acknowledge that these are legitimate concerns, and show that these are just two of several problems with compositional models.

COMPOSITIONAL CONCEPT MODELS

Some authors speak of SNOMED concept composition as though it is a single well-defined model. However, in examining the literature on SNOMED and looking at different interpretations and implementations, we have identified at least three major variations in how compositional concepts have been represented in the past. In order to more fully understand these approaches to compositional concept representation using SNOMED, we name and describe them, along with their strengths and weaknesses, and also describe our current approach based on SNOMED RT [4]. In this paper, the four approaches are called Compositional Concept Models (CCM) 1 through 4:
CCM-1: Unconstrained composition
CCM-2: Multi-axial composition
CCM-3: Attribute-value composition
CCM-4: Foundational model composition

CCM-1: Unconstrained Composition

CCM-1 might be called "unconstrained" concept composition. The basic 'idea is that elementary or atomic concepts are enumerated and classified in a nomenclature, and then a compositional concept can be constructed by combining more than one atomic concept. Even though the atomic concepts may be from different "axes," in CCM-1 there are no significant constraints on how the combination is to take place; the structure is simple concatenation. Interpretation of the meaning of the concatenated string usually is dependent on the knowledge of the individual who examines the string; computer-based interpretation of such compositional concepts is fraught with ambiguities and duplications. CCM-1 has been criticized by many authors. Two of the main criticisms are: 1) that a given concept can be represented many different ways, and 2) that it is not possible for the computer to recognize the equivalance of these different ways of representing the concept.

Many of the studies of SNOMED's expressiveness seem to have assumed an unconstrained (CCM-1) compositional model [2]; it is possible, but has not been determined whether a more constrained or principled model of composition would have resulted in less expressiveness in these studies.

CCM-2: Multi-axial Composition

CCM-2 might be called "multi-axial" composition. This model was described in detail in the SNOMED II Coding Manual of 1979 [5]. The essence of the model is that there is a set of "axes" that can be combined to form composite concept representations or assertions. CCM-2 and CCM-1 are not often differentiated; however, CCM-2 is much stronger semantically, and also has less expressive flexibility.

This model has a long history in coding systems. The original SNOMED, published in 1976-77, was based in part on SNOP (1965) and SNDO, both of which had a "multi-axial" nature. For example, anatomic site (T for topography) and structural change (M for morphology) are separately enumerated (in SNOP, ICD-O, and SNOMED), and their basic elements can be combined. Thus, instead of having a separate code for every possible tumor morphology in every possible anatomic location, one simply combines the morphology with the topography. For example, adenocarcinoma of the stomach would be coded as a combination of the M-code for adenocarcinoma with the T-code for stomach.

The axes in SNOMED II are Procedure, Topography, Morphology,, Etiology, Function, and Disease. Each axis is given a single field in the coding table, and three additional fields are added to the table: the "context," represented by Information Qualifiers (IQ), the time (duration), and finally linkages to other concepts. Each individual assertion or concept is represented as a row in a table, and combined assertions can be represented by linking successive rows together.

Figure 1a shows the SNOMED II coding template, and Figure 1b shows the representation of a composite concept using this template.

IQ	P	T	M	E	F	D	TIME	LINK

Figure 1a: SNOMED II Coding Template

IQ	P	T	M	E	F	D	TIME	LINK
(FD) FINAL DIAGNOSIS		T-64300 DUODENUM	M-34000 OBSTRUCTION , NOS					(DT) DUE TO
(FD) FINAL DIAGNOSIS		T-58700 AMPULLA OF VATER	M-81403 ADENO-CARCINOMA					

Figure 1b: SNOMED II Coding Template showing the codes for a final diagnosis of duodenal obstruction due to adenocarcinoma of the ampulla of Vater.

CCM-3: Attribute-value Composition

CCM-3 might be called "attribute-value" composition. The need for explicit attributes, instead of simply a list of a few axes, was apparent to the authors of the SNOMED II coding manual. Where CCM-2 conflated the axis and the attribute, CCM-3 splits them out, as in the example in figures 2a and 2b, which represent the concept "gunshot wound of forehead, by handgun, with hypovolemic shock, homicide."

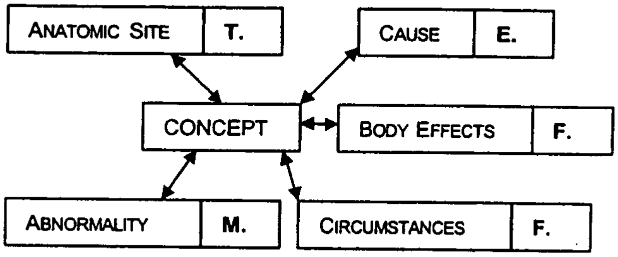


Figure 2a: Basic multi-axial concept template (from SNOMED II Coding Manual, Figure 6)

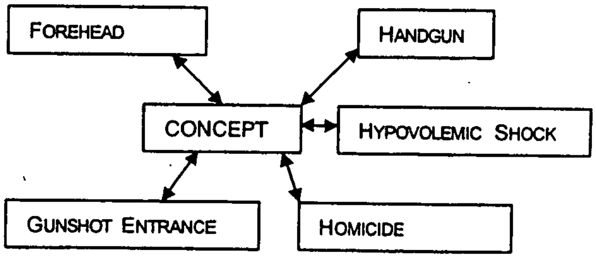


Figure 2b: Template from 2a with specific examples (from SNOMED II Coding Manual, Figure 8)

It is clear from these two figures that we have an example of a single concept that requires two codes

It has been shown that computing subsumption for expressive DLs is computationally intractable; in order to achieve tractable (worst-case) subsumption, the concept-forming operators used in the DL must be restricted.

Horrocks describes an experiment in which he attempts to compare the GRAIL language with LOOM, another description logic language, and he describes the concept-forming operators that are used. [7] CCM-4 is based on a tractable description logic which uses the same set of concept-forming operators (K-REP). [9] In Table 1, commonly-applied concept-forming operators are listed, and the concept-forming operators used by CCM-4 and by GRAIL are listed with an asterisk.

	Operator name	Notation
1*	Top (everything)	$\top$
2*	Bottom (∅)	$\perp$
3*	Conjunction	$C_1 \sqcap \ldots \sqcap C_n$
4*	Exists restriction	$\exists R.C$
5.	All restriction	$\forall R.C$
6.	Disjunction	$C_1 \sqcup \ldots \sqcup C_n$
7.	Negation	$\neg C$
8.	Number restriction	$\geq nR.C$
9.	Number restriction	$\leq n\mathrm{R}.C$

Table 1. General concept-forming operators in description logics. *=operators used in GRAIL and in CCM-4.

We can now show how to solve the "acute appendicitis" example, commonly used as an example of the inadequacies of CCM-1 and CCM-2. Table 2 shows the relevant SNOMED codes and terms.

D5-46210	Acute appendicitis, NOS
D5-46100	Appendicitis, NOS
G-A231	Acute
M-40000	Inflammation, NOS
M-41000	Acute inflammation, NOS
T-59200	Appendix, NOS
G-C006	Has location (In)

Table 2: SNOMED Codes related to the concept "Acute appendicitis"

Tables 3 through 6 show the different possible representations of acute appendicitis using each of the compositional concept models described in this paper.

D5-46210	Acute appendicitis
G-A231,D5-46100	Acute + appendicitis
M-41000, G-C006, T-59200	Acute inflammation + In + appendix
G-A231, M-40000, G-C006, T-59200	Acute + inflammation + In + appendix

Table 3: Acute appendicitis represented by CCM-1

IQ	P	T	M	E	F	D	T	L
						D5-46210 Acute Appendicitis		
G-A231 Acute						D5-46100 Appendicitis		
		T-59200 Appendix	M-41000 Acute Inflammation					
G-A231 Acute		T-59200 Appendix	M-40000 Inflammation					

Table 4: Acute appendicitis represented by CCM-2

D5-46210		
D5-46100	has-course	G-A231
M-41000	assoc-topography	T-59200
M-40000	has-course	G-A231
	assoc-topography	T-59200

Table 5: Acute appendicitis represented by CCM-3

D5-46210		
D5-46100	has-course	G-A231
DF-00000	assoc-topography	T-59200
	assoc-morphology	M-41000
DF-00000	has-course	G-A231
	assoc-topography	T-59200
	assoc-morphology	M-40000

Table 6: Acute appendicitis represented by CCM-4

Note that because CCM-4 has a foundational model of disease that links all disease expressions to a "root" concept of disease (DF-00000), the last two CCM-3 representations are explicitly invalid as representations of disease.

from the same axis (F). Note that such a representation cannot properly be made using the SNOMED II Coding Template (CCM-2). Figure 2a explicitly lists the attributes, while figure 2b lists the values, and thus the concept may be represented as:
Concept =

Anatomic site: Forehead
Abnormality: Gunshot entrance
Cause: Handgun
Circumstances: Homicide
Body Effects: Hypovolemic shock

Examining this representation, it is clear that each line has two elements, the first of which describes an attribute and the second of which describes the value of that attribute. In the terminology of description logic, the first element may be called the "defining relationship" and the second element may be called the "value restriction.." [6] In the example, "circumstances" could be regarded as a defining relationship, and "homicide" a value restriction.

One way of interpreting CCM-2 that helps to understand the origin of its basic weakness is that it enumerates and classifies the values, but leaves the attributes implicit (in the axis) or undefined. In order for there to be clear and unambiguous representation of concepts, and consistent use of the terminology, each compositional concept expression should explicitly give both the attributes and their values. This is what CCM-3 does that CCM-2 does not do.

The total number of possible attributes is much larger than the number of "axes," or orthogonal semantic groupings of basic concepts. Conflating the attributes with the axes results in limited expressiveness. Thus CCM-3 has more flexibility and expressiveness than CCM-2, and also has a stronger explicit semantics. CCM-3 corresponds quite closely to the NHS Clinical Terms compositional model, as described in the "Version 3.1 File Structure: Qualifier Extensions".

The SNOMED III G axis (general linkage and modifiers) contains a relatively large collection of terms that might be used as "attributes" in an attribute-value type of compositional concept representation. However, particular concepts are not explicitly identified as attributes, and no coherent effort was made to identify a set of such attributes (or defining relationships) that, taken together, could form the basis for a "foundational model" of composite concept representation. Thus CCM-3 lacks a consistent set of defining relationships as part of the compositional model. CCM-4 addresses this deficiency.

CCM-4: Foundational Models, Description Logics

CCM-4 may be called "foundational model" composition. [1] It builds upon CCM-3 by adding a formal semantics based on description logics, and by explicitly requiring a foundational set of defining relationships. The difference between CCM-3 and CCM-4 is therefore primarily not in the format of the representation of composite concepts, but in the models underlying the format.

Foundational models

Foundational models consist of sets of defining relationships, along with the assumptions and conceptual model that underlies these relationships. For example, the SNOMED multi-axial coding of tumors can be re-cast into CCM-4 by declaring the defining relationships "has-topography" and "has-morphology" as the core of the foundational model for compositional encoding of tumors.

Description logics

Description logics are formal subsets of predicate logic which typically have a formal semantics based on Tarski-style denotational semantics. [7] A more comprehensive review of the characteristics of description logics is beyond the scope of this paper, and can be found elsewhere. [8] However, a few basic definitions should suffice to describe the importance of description logic as a foundation for a compositional model of medical concepts.

Description logic statements are used to denote the essential characteristics of concepts; that is, to create a formal representation of the semantic definition of a concept, based on those features or characteristics that are always true and that differentiate one concept from another.

Description logic statements can be composed of "concept-forming operators," that is, operators that take concepts and defining relationships (also known as "roles") and combine them to form new logical expressions that define the meaning of a concept.

A description logic "engine" reads DL statements and then computes subsumption relationships between concepts; in other words, the DL engine can tell from the DL definition of two terms whether one is a specialization or generalization of the other.

TERMINOLOGY CONVERGENCE

NHS Clinical Terms version 3 adopts a object-attribute-value triple approach to representing concepts. [11] Thus it explicitly identifies attributes, which may be interpreted as "defining relationships." In fact, these are used as the basis for "semantic definitions."

The same tables used to express semantic definitions are also used to create templates for composition of concepts. Each term that can be modified is listed with the defining attributes that may modify it, and the values (or value sets, by reference) that may be used. Convergence would be possible based on harmonization of the attributes in these template files with the SNOMED foundational models.

GRAIL uses the same concept-forming operators as CCM-4. [7] This limited set of concept-forming operators seems to be satisfactory for a significant part of terminological representation in health and medicine. In addition, GRAIL has three kinds of sanctioning: 1) conceivable, 2) grammatical, and 3) sensible. [10] CCM-4 provides conceivable sanctioning through creation of defining roles and value restrictions. Sensible sanctioning is required primarily for generative terminology, which in turn is mainly a user-interface issue rather than a reference terminology issue.

Further examination of the GRAIL CORE model, and the intermediate representation being used for clinical modelers, may allow consideration of further convergence of it and SNOMED's foundational models.

CONCLUSION

Reliable and accurate compositional concept representation is now feasible through the combined use of a reference terminology, description logic semantics, and a set of foundational models. Implementation of this approach in the Kaiser-Permanente Convergent Medical Terminology project will afford ample opportunity to evaluate its effectiveness.

Acknowledgments

This work is supported in part by a grant from the College of American Pathologists, and in part by Kaiser Permanente.

References

1. Campbell KE, Das AK, Musen MA. A Logical foundation for representation of clinical data. J Am Med Informatics Assoc 1:218-232, 1994.

2. Cote RA, Rothwell DJ, Palotay JL, Beckett RS, Brochu L, eds. The Systematized Nomenclature of Human and Veterinary Medicine: SNOMED International. Northfield, IL: College of American Pathologists, 1993.

3. Chute CG, et al. The content coverage of clinical classifications. JAMIA 3:224-233, 1996.

4. Spackman KA, Campbell KE, Cote RA. SNOMED RT: A reference terminology for health care. In: AMIA Annual Fall Symposium, 640-644, 1997.

5. Gantner GE, Côté RA, Beckett RS, eds. Systematized Nomenclature of Medicine, Coding Manual. Skokie, IL: College of American Pathologists, 1979.

6. Dolin RH, et al. Evaluation of a "lexically assign, logically refine" strategy for semi-automated integration of overlapping terminologies. JAMIA 5:203-213, 1998.

7. Horrocks IR. A comparison of two terminological knowledge representation systems. University of Manchester thesis. July 1995.

8. Woods WA, Schmolze JG. The KL-ONE family. Computers and Mathematics with Applications -- Special Issue on Artificial Intelligence, 23(2-5):133-177, 1992.

9. Mays E, Dionne R, Weida R. K-REP system overview. SIGART Bulletin 2(3):88-92, 1991.

10. Rector AL, Bechhofer S, Goble CA, et al. The GRAIL concept modelling language for medical terminology. *Artificial Intelligence in Medicine* 9(2):139-171, 1997.

11. NHS Centre for Coding and Classification. Read Codes File Structure Version 3.1 - The Qualifier Extensions. January 1995.

12. Evans DA, et al. Toward a medical concept representation language. JAMIA 1:207, 1994.

PubMed Central
· Journal List
· Search
· Utilities

JAMIA
The Journal of the American Medical Informatics Association
Subscribe | AMIA Home | Join AMIA | Search AMIA

Exhibit D

PubMed Central
▸ Abstract
■ Full Text
▸ Figures/Tables
▸ PDF
▸ Contents
▸ Archive

PubMed
Articles by:
▸White, M.
▸Kolar, L.
▸Steindel, S.

and links to:
Related articles
Show

J Am Med Inform Assoc. 1999 May; 6(3): 185–194.

Evaluation of Vocabularies for Electronic Laboratory Reporting to Public Health Agencies

Mark D. White, Linda M. Kolar, DVM, MPH, and Steven J. Steindel, PhD

Centers for Disease Control and Prevention, Atlanta, Georgia.

Correspondence and reprints: Steven J. Steindel, PhD, Centers for Disease Control and Prevention, 4770 Buford Highway, NE (MS G-23), Atlanta, GA 30341. e-mail: <sns6@cdc.gov >.

Received 1998 August 11; Accepted 1999 January 12.

This article has been cited by other articles in PMC.

Abstract

Clinical laboratories and clinicians transmit certain laboratory test results to public health agencies as required by state or local law. Most of these surveillance data are currently received by conventional mail or facsimile transmission. The Centers for Disease Control and Prevention (CDC), Council of State and Territorial Epidemiologists, and Association of Public Health Laboratories are preparing to implement surveillance systems that will use existing laboratory information systems to transmit electronic laboratory results to appropriate public health agencies. The authors anticipate that this will improve the reporting efficiency for these laboratories, reduce manual data entry, and greatly increase the timeliness and utility of the data. The vocabulary and messaging standards used should encourage participation in these new electronic reporting systems by minimizing the cost and inconvenience to laboratories while providing for accurate and complete communication of needed data. This article describes public health data requirements and the influence of vocabulary and messaging standards on implementation.

Introduction

Widespread use of clinical laboratory information systems (LISs) and development of electronic data interchange standards provide the nation's public health agencies with the opportunity to supplement or replace current mechanisms for reporting data[1] with far more efficient and timely electronic systems. The current paper systems are inefficient for participating laboratories and limit use of these data by public health agencies. In particular, minimizing the delay between completion of reportable laboratory test results and transmission of these data to public health agencies would improve our ability to rapidly detect and identify events of public health significance, including outbreaks of infectious diseases and changes in antimicrobial resistance patterns. We discuss here one of the key steps toward implementing electronic reporting systems: the evaluation of common nomenclatures and data transmission standards.

These standards must be sufficiently flexible and comprehensive to accommodate the broad and varied reporting needs of both the laboratory community and public health. The types of health care facilities that transmit laboratory data are diverse, ranging from small hospitals and public health clinics to large national reference laboratories. The business needs of these facilities and the information systems that support them are also diverse. To permit nationwide participation, it will be necessary to select one or more vocabularies that enable most reporting laboratories to accurately translate data from their existing LIS.

The basic data requirements for these reportable laboratory test results include such items as test or procedure identifier, specimen type, anatomic collection site (when relevant), reference range, unit of measure, result, and patient demographics. Demographic information is of particular importance for positive results associated with infectious organisms, environmental agents, or patients with repetitive events. In such cases, both demographic and clinical information is often needed in order to understand the source of the problem and formulate strategies for preventing further transmission or exposure. These data must be timely, accurate, and accessible to meet the needs of the agencies responsible for monitoring and protecting the public's health. Identifying infectious disease outbreaks and environmental hazards are but two examples of public health activities that rely on the timeliness, quality, and accessibility of surveillance data.

Effective surveillance often relies on both the test results reported by the laboratory and clinical data provided by the physician. Public health agencies do not all require the same level of detail from the information derived from surveillance data (e.g., local investigation of an infectious disease outbreak vs. national breast cancer rates). For accommodating the data needs of the various public health agencies conducting disease surveillance, a vocabulary capable of describing concepts in varying levels of detail is required. Ideally, the subset of terms selected for reporting laboratory results should be part of a more comprehensive multipurpose vocabulary that is also capable of describing clinical data. This would permit clinical data from physicians to be included in reports without requiring health care facilities to support yet another specialized vocabulary.

This paper provides a brief overview of public health needs and data security concerns, followed by more in-depth discussions of vocabulary characteristics and their potential impact on the ability of laboratories to implement electronic reporting of test results to public health agencies. The choice of a particular vocabulary may profoundly affect a laboratory's ability to participate in this type of reporting and influence the laboratory's accuracy in mapping terms. Representation of concepts, concept identifiers, and vocabulary evolution are addressed with regard to the needs of public health agencies. Also included is a preliminary assessment of the ability of U.S. LIS vendors to comply with the current recommendations of the Centers for Disease Control and Prevention (CDC) for electronic laboratory reporting.

▸ Top
▸ Abstract
▸ Introduction
Public Health

Public Health Reporting Characteristics

Public health reporting needs represent both a subset and an extension of

- Reporting Characteristics
- Security Concerns
- Vocabulary Characteristics
- Implementation
- Conclusions
- References

requirements for the computerized patient records. They include a variety of infectious diseases, chronic diseases such as lung cancer, and environmentally induced conditions such as lead toxicity. In many cases, the reports may contain epidemiologic information such as risk factors (e.g., smoking history for lung cancer or age of housing for lead reporting).

Public health agencies receive reports from both clinical and nonclinical sources. Information on trauma or accidents may be received from an emergency department, police department, or workplace. Other reports may provide data on the quality of the water supply from the local (public or private) water treatment plant or beach. Information reported on animal or insect vectors can be used to trace the sources of infectious disease outbreaks. Results from tests of soil, air, or building samples are used to help locate the source of environmentally induced outbreaks.

Describing public health concepts through defined vocabularies requires that the above examples, and many similar ones, be considered. Fortunately, the actual information that the vocabularies need to express for electronic laboratory reporting to public health agencies is but a small subset of the information that would be required for a computerized patient record. As we shall describe, information models for reporting public health data are being adapted to meet electronic reporting requirements. In choosing vocabularies for public health reporting, we must consider the diversity of data required and the sometimes rapidly changing informational content of the data. Within this construct, public health agencies need to develop a system that is compatible with the information systems found in the private sector, from which much of the data will be derived.

- Top
- Abstract
- Introduction
- Public Health Reporting Characteristics
- Security Concerns
- Vocabulary Characteristics
- Implementation
- Conclusions
- References

Security Concerns

The CDC recognizes that the issue of data and transmission security is paramount to the implementation of any electronic public health reporting system and is developing both methods and procedures to address this issue. Programs used by CDC have transmitted sensitive information electronically using encryption and dial-up modems for many years. Two examples of such programs are the Laboratory Information Tracking System (National Center for Infectious Diseases) and the HIV/AIDS Reporting System (National Center for HIV, STD and TB Prevention), both widely used by public health departments. Recently implemented was a policy document applying to all`programs sponsored by CDC that, in principle, allows sensitive data transmission via Internet using an appropriate encryption method of at least 128 bits, message authentication, and message nonrepudiation (CDC Internet Standard 98.1). In process is the establishment of a secure data facility consisting of a certification server and a public/private key encryption system with digital signatures based on the X.509 standard (ITU-T X.509, version 3). A similar system is used by New York State for the Internet transmission of sensitive data (Ivan Gotham, NY State Health Department, private communication). Lastly, we are engaged in extensive policy discussions with our public health partners concerning the decisions and agreements that must be in place before any transmission system can be widely implemented. Further discussion of this important issue is beyond the scope of this article.[2]

- Top
- Abstract

Vocabulary Characteristics

The concepts that a laboratory system can express and report are controlled not only by the richness of a vocabulary's content but also by its structure. The manner in which a vocabulary combines concepts, discriminates between terms, and evolves will affect the clarity, accuracy, flexibility, and level of detail with which it may express the important clinical and laboratory data needed for public health surveillance.

Representation of Concepts

A vocabulary term can represent a single atomic concept or an aggregate concept. For example, ***Mycobacterium tuberculosis*** is a single atomic concept, whereas ***Mycobacterium tuberculosis*** detected in **sputum** by **DNA probe** is an aggregate concept consisting of three atomic concepts (organism + specimen type + testing methodology). Precoordinated vocabularies assign unique concept identifiers (codes) to predefined aggregate concepts (terms) and can therefore convey complex information without ambiguity.[3] This eliminates nonsensical or undesirable combinations of individual concepts, since the aggregate concepts are precoded in the only manner permitted. For example, the contemporary precoordinated vocabulary known as the Logical Observation Identifier Names and Codes (LOINC) uses the concept identifier 5289-4 to represent an aggregate concept (term) that consists of six individual concepts (Table 1). This fully specified term explicitly describes a reagin antibody analyte concentration conducted on a cerebrospinal fluid specimen collected at a single point in time and measured quantitatively by a flocculation method. Such precoordinated vocabularies are an unambiguous and precise means of representing detailed information as a simple code. This clarity, however, does not come without tradeoffs.

In practice these terms are so highly specific and complex that laboratory staff may not have sufficient expertise or knowledge to successfully translate tests to LOINC.[4] Where such expertise does exist, disagreements among local experts can still introduce variance in how identical tests are coded at different institutions. Delays in development or distribution of an updated version of the vocabulary may result in some laboratories temporarily reverting to paper reporting systems or mapping new aggregate concepts to existing but inaccurate terms. Public health agencies often need to aggregate the data received from reporting laboratories and will need laboratories to code reportable information accurately and consistently. If some laboratories map to the fully specified codes while others map using a "best fit" technique, then the data cannot be combined or compared and lose much of their utility.

Precoordination of fully specified terms also requires that assumptions be made about health care business practices and data structures. In facilities that deviate from the assumed model, it may be difficult or impossible to map the individual concepts embedded in the aggregate concept. Continuing with the example above, laboratory systems may not define each test in terms of mass concentration (MCNC), substance concentration (SCNC), or analyte concentration (ACNC). Likewise, many LISs may not be able to discriminate on the basis of scale type, methodology, or time. Some may not even store specimen type in the LIS, relying instead on the specimen type identified in the collection order. In such cases, substantial effort and resources may be required to add the new fields and data necessary to accurately map all the individual concepts contained in the precoordinated terms. Developing and maintaining complex cross-references may also be required. As a result, the reporting facility could conclude that it is more accurate or cost effective to maintain the current paper reporting system.

Since terms must be provided for all the potentially useful combinations of individual concepts, these concepts occur redundantly throughout the vocabulary. Precoordinated vocabularies are therefore large relative to the number of unique concepts that they describe. If such a vocabulary contains 250 terms representing all the potentially useful combinations of test name and sample type for a specific testing methodology, and if a new or revised testing methodology is subsequently developed, it may be necessary to create as many as 250 new terms. In addition, the original terms must be maintained to preserve the integrity of previously coded data, permit laboratories a gradual transition to the new methodology, and accommodate those who continue using the old method. It is uncertain how well a precoordinated vocabulary can handle the inevitable combinatorial explosion of tests, methodologies, specimens, and results.

Postcoordinated vocabularies, in contrast, assign codes to individual concepts that the laboratory system can use to compose any number of aggregate concepts.[5] For example, six individual concept codes could be combined to compose an aggregate concept equivalent to the fully specified term from the example given in Table 1 (LOINC term 5289-4). Such a vocabulary is generally more flexible in representing new or revised concepts with existing codes. It is also less redundant and thus potentially smaller. Laboratory systems need only map to the individual concepts that fit their business practices and data structures. The Systematized Nomenclature of Human and Veterinary Medicine (SNOMED)[6] is an example of a contemporary postcoordinated vocabulary. When new aggregate concepts are required, the laboratory system may often be able to combine existing individual concept codes to compose the new terms. Similarly, developing a new or revised testing methodology, as in the example above, would require adding a single code that would then be combined with existing test name codes and specimen codes to represent the new term, rather than adding 250 new terms. Since terms consist of discrete (individually) coded concepts, it is also a relatively simple matter to parse them into their constituent concepts. This in turn provides greater flexibility for sorting and querying the data in ways that are meaningful for the various public health agencies. While precoordinated terms can also be disassembled, the resulting individual concepts may not translate to those commonly used in other vocabularies and may not be suitable for complex queries.

Vast expressive power, concept richness, and flexibility also do not come without tradeoffs. Since postcoordinated vocabularies often allow aggregate concepts to be coded in more than one way, it may be necessary to develop guidelines that identify preferred combinations and discourage the use of undesirable or nonsensical terms. Several agencies currently meet on a regular basis to discuss common concepts, methods, and terminology for use in public health (Table 2). These agencies could be asked to develop and distribute appropriate guidelines for creating new public health terms from existing concepts. Until such compositional guidelines can be developed and distributed, public health agencies must carefully weigh the expressive nature of postcoordinated vocabularies against the potential for users to transmit undesirable combinations of concepts.

In summary, the unambiguous nature of terms in a precoordinated vocabulary offers the clarity and precision of reporting that public health agencies need to aggregate and analyze the data, but such terms may be difficult or impossible for laboratories to integrate with their current information systems and are more likely to be negatively affected by the inevitable combinatorial explosion of terms. The vast expressive power, concept richness, and flexibility of a postcoordinated vocabulary seem better suited to public health surveillance requirements and the diverse information system capabilities of laboratories, but the postcoordinated vocabulary may require additional guidelines for the composition of appropriate and consistent terms.

Hierarchic versus Context-free Concept Identifiers

Hierarchic concept identifiers are codes that indicate the ordered position of a concept in the vocabulary.[7] For example, the SNOMED hierarchic identifier for *Escherichia coli*, serotype O157:H7, is L-15611. The "L" in the concept identifier indicates that the code represents a living organism. Concept identifiers beginning with "L1" describe Bacteria and *Rickettsiae.* Entries beginning with "L15" are *Enterobacteriaceae.* Entries beginning with "L156" describe members of the genus *Escherichia.* Thus, "L-15601" identifies *Escherichia coli* and "L-15611" identifies *Escherichia coli*, serotype O157:H7. This offers users a means of understanding the relationships and differences between concepts and can enhance the ability to map terms accurately. A disadvantage of vocabularies using these hierarchic identifiers is that often only a finite number of terms can be added within each level of the hierarchy, and so expansion is limited.[8] Reclassification of coded concepts can also be problematic, since it may require changes to both the concept and its identifier.

Vocabularies that use context-free concept identifiers are preferred to those that use hierarchic concept identifiers, because context-free identifiers do not restrict the number of terms that can be added and permit additional flexibility in reclassifying terms.[9]

Vocabularies that separate the hierarchic structure from the concept identifier offer public health and laboratory communities the best alternative. These vocabularies still assist users in accurately discriminating between concepts while using context-free identifiers that do not restrict the addition and reclassification of terms.

Vocabulary Evolution[10]

A controlled vocabulary used for reporting laboratory results to public health agencies must be able to maintain order and integrity as it evolves, i.e., the rules governing change must be applied in a consistent manner and the vocabulary must retain its compatibility with previous versions. Many changes will occur in the identification and description of etiologic agents and in the public health requirements for reporting of disease. Our expanding knowledge base will cause researchers to reorder classification schemes, rename known agents, and allow them to identify agents of both new diseases and diseases currently described as being "of unknown origin." Laboratory methods will continue to be developed, refined, and discarded. Public health agencies will revise reportable disease and event lists to reflect these changes and to meet tomorrow's challenges. It is thus vital that the vocabulary used for reporting laboratory test results be capable of evolving at a rate sufficient to meet these needs.

Yet these changes must be made in a careful, well-documented fashion. Users must not only be aware of additions to, deletions from, and name changes in the vocabulary, but also be cognizant of the reasons for the changes and the impact they may have on applying codes in the future and on interpreting old and new data. The creators of the vocabulary should provide both a formal syntax of changes, to convey the surface differences, and a semantics of the changes, to describe how the meaning of a term is or is not altered during the process.[11] This will ensure that the vocabulary evolves in a logical manner and will allow backward and forward compatibility in collected data. In addition, new releases of the vocabulary must be clearly distinguished, so that users can identify which version they are using and can track by date the changes that are made. Without documented evidence of controlled evolution, it will be difficult to combine data from various sources with assurances of compatibility and data comparability.

LOINC was introduced in 1995 and experienced an initial period of rapid growth. Since then updates have been distributed via Internet or diskette about twice a year. LOINC is distributed free of charge and is maintained through grant support. The Regenstrief Institute and the LOINC Committee have indicated that they will maintain the database while grant support is available (at least until October 1999). Snomed was introduced in 1976 and has been adding new concepts as required and distributing them as part of an annual update on diskette or compact disc. Snomed is distributed to licensed users and is professionally maintained through license fees.

- Top
- Abstract
- Introduction
- Public Health Reporting Characteristics
- Security Concerns
- Vocabulary Characteristics
- Implementation
- Conclusions
- References

Implementation

To take advantage of the potential benefits of electronic laboratory reporting, CDC, in consultation with its partners, has elected to evaluate Health Level Seven (HL7), version 2.3,[12] as the messaging standard for pilot testing the transmission of reportable laboratory test results to public health agencies.[13] These pilot studies will assist in identifying the obstacles and problems that must be overcome before widespread deployment. As part of this test implementation, these public health agencies will use the unsolicited transmission of an observation (ORU) transaction set. The HL7 v2.3 documentation strongly encourages the use of universal identifiers in the observation identifier (OBX-3) segment of this transaction set.[*] It also specifies LOINC as one of the possible universal identifiers that could be used in this segment. Based on this documentation and suggestions from HL7 members, CDC has chosen to evaluate the use of LOINC as a universal identifier in OBX-3 during our HL7 pilot studies. To minimize difficulties in aggregation for observation (result) values that are non-numeric, such as organism names, we will require laboratories participating in the pilot studies to use a coded element (SNOMED codes) rather than text in OBX-5.

CDC is currently considering several HL7-related activities that will allow the agency to evaluate and develop the potential to use LIS to transmit HL7 messages of results of laboratory tests for infectious diseases or detection of incident cancer cases. Such studies will provide valuable information for assessing the future applicability of direct reporting to appropriate public health agencies. Among the proposed activities are a survey of U.S. LIS vendors to determine their ability to implement electronic laboratory reporting of clinical and anatomic laboratory data to public health agencies using the HL7 messaging standards, and development of an implementation specification based on the CDC's HL7 electronic laboratory reporting message recommendations and several pilot projects. These pilot studies will provide an assessment of the effectiveness of both the implementation specification and the transmission system and will assist public health agencies in developing a national electronic laboratory reporting system. To optimize our data collection efforts, CDC has been conducting a nationwide retrospective survey to gather objective information regarding the type and volume of laboratory testing that was performed at representative testing locations in 1996.[14]

The pilot studies will evaluate the ability of laboratory information systems to use fully specified and precoordinated observation identifiers in OBX-3. They will also identify and evaluate the differences between the business practices and data structures of laboratories and the model used to precoordinate the LOINC terms.

Laboratory Information Systems

Large U.S. laboratories often have adequate resources and technical expertise to program their own information systems. These laboratories should be capable of implementing the vocabularies that public health agencies recommend for electronic laboratory reporting. However, many U.S. laboratories rely on LIS vendors to provide the software, hardware, and programming to meet their information system needs, and our recommendation may have a much greater impact on their ability to participate. To gauge the ability of these laboratories to participate in our pilot studies, we conducted unstructured telephone interviews with 11 HL7-capable LIS vendors who agreed to participate in this informal survey. These vendors represent approximately 52 percent of hospital laboratory, 25 percent of independent laboratory, 29 percent of clinic or group practice laboratory, and 56 percent of other LIS installations as reported by 67 LIS vendors in a 1995 review.[15]

These companies were experienced in traditional electronic data interchange relationships, where they define the meaning of every test code with each partner with whom they exchange information. They were generally less familiar with the differences between precoordinated and postcoordinated vocabularies. All the vendors welcomed the concept of a universal test identifier. While 10 of the 11 vendors were aware of the existence of the LOINC codes, only two were aware that each LOINC code incorporates concepts other than test or procedure name. These two were also the only vendors who indicated they currently store the information required to complement all the individual concepts (dimensions) of a fully specified LOINC code (Table 3). Of the remaining vendor systems that included fields that could be mapped to two or more dimensions of the LOINC code, these fields occurred in two or more tables in their systems. Most indicated that fields identifying specimen type are located in separate tables from test identifications or test descriptions. Nine vendors believed that the additional fields and complex cross-references necessary to implement fully specified observation identifiers such as LOINC codes would be both difficult and expensive. These nine vendors also indicated that they could probably provide laboratories with the ability to transmit reportable test results to public health agencies in the near future if they used a less complex mapping for OBX-3, such as that permitted by a postcoordinated vocabulary or a simplified universal test identifier.

Simplified Observation (Test) Identifiers

In 1997, a group of state epidemiologists and public health officials developed a preliminary table containing approximately 64 reportable entities (infectious organisms and agents) important to public health surveillance and the accepted testing methodologies and procedures used in their identification.[13] Members of the group used a precoordinated vocabulary (LOINC) to provide coded terms for the preferred tests and procedures needed to identify these reportable entities. The group was able to describe most of the 64 reportable entities using 280 LOINC codes.

Codes were not yet available for some entities. To permit useful data aggregation and analysis, many of these would require a coded element in OBX-5 for non-numeric results such as the name of the organism identified.

As one means of providing a simplified observation identifier, we attempted to represent the 64 reportable entities, methods, and procedures using existing SNOMED codes (v3.3) from the procedure, living organism, and modifier axes. For this exercise, we chose to use a two-component code, but more complex and specific terms can be composed. This would allow a laboratory system to compose as detailed a term in OBX-3 as their data structures permit without imposing the limitations of fully specified precoordinated terms. We found we could represent most of these concepts with 50 procedure codes, a living organism code for each reportable entity, and 2 modifier codes (121 codes total). Adding approximately a dozen coded concepts (2 living organism codes and 10 procedure codes) would permit all the entities to be identified and would increase the specificity of the coded methodologies. The resulting terms were less detailed than the LOINC codes, containing only information on the test procedure or method and organism identified.[16] However, additional details can be placed elsewhere in the reporting message. For example, the HL7 ORU transaction set contains an observation request (OBR) segment that provides separate fields for the specimen source and anatomic site, and an observation result (OBX) segment that contains fields for units, reference range, and further distinction of the methodology. It should not be necessary to also include this information in the laboratory observation (test) identifier (OBX-3) for public health reporting. Use of a simplified observation identifier in conjunction with the information in these fields can provide all the information required for reportable laboratory test results without the potential mapping limitations that may be associated with a fully specified and precoordinated observation identifier.

The following examples illustrate the differences between using fully specified LOINC codes containing up to six dimensions and simplified two-dimensional SNOMED terms as observation identifiers in OBX-3. When the observation values (results) are non-numeric, such as for organism identification, a SNOMED code is used in OBX-5 to supplement the LOINC code and provide a coded entry for data aggregation and analysis. The SNOMED examples use a two-dimensional observation identifier composed of a procedure code and a living organism code for OBX-3. A living organism code is also used in OBX-5. This may appear redundant in some cases, but where the observation value in OBX-5 is qualitative, semiqualitative, or numeric, it is at times necessary to include the organism name in the observation identifier. For example, when *Yersinia pestis* (the plague) is reported, the preferred antibody test is the enzyme-linked immunosorbant assay (ELISA). A reportable result for this assay would be numeric (⩾1:64). In this case we would use the codes P3-70200 | L-1E401 in OBX-3 to represent ELISA for *Yersinia pestis* and place the numeric result (titer) in OBX-5. The following examples utilize the specimen source information coded in the OBR-15 segment, which is required regardless of the vocabulary used in OBX-3.

Example 1: Reportable Results Data To Be Coded

Rabies virus identified by antibody neutralization in serum or cerebrospinal fluid

Rabies virus identified by direct fluorescent antigen detection in tissue

> or other (unspecified specimens)
>
> Rabies viral culture in saliva, cerebrospinal fluid, central nervous system, tissue, or other (unspecified) tissue
>
> LOINC can represent this information with eight multidimensional codes (Table 4).

Snomed can represent this information by combining a procedure code with a living organism code in OBX-3. The specimen source information is obtained from the OBR segment (OBR-15), where it can be coded using the standard HL7 specimen table (0070) or terms from the SNOMED topography table (Table 5).

> Example 2: Data To Be Coded
>
> Microbial culture of *Streptococcus pyogenes* in blood, cerebrospinal fluid, pleural fluid, peritoneal fluid/ascites, wound, or other

LOINC can describe most of these microbial cultures with six multidimensional codes. For representing the organism that was identified (*Streptococcus pyogenes*), a coded element (SNOMED) code is required in OBX-5 to permit data aggregation and analysis (Table 4).

Snomed can represent the same organism identification by combining a procedure code (Microbial Culture) and a living organism code (*Streptococcus pyogenes*) in OBX-3. The living organism code (*Streptococcus pyogenes*) is also used in OBX-5. The specimen source information is coded in the OBR segment as above (Table 5).

▸ Top
▸ Abstract
▸ Introduction
▸ Public Health Reporting Characteristics
▸ Security Concerns
▸ Vocabulary Characteristics
▸ Implementation
■ Conclusions
▸ References

Conclusions

For widespread implementation of electronic laboratory reporting, public health agencies must first ensure that the electronic transmission, storage, and use of this information is at least as confidential and secure as current systems. Next we must ensure the most complete surveillance data possible by selecting coding schemes, vocabularies, and messaging standards that allow reporting laboratories to participate and accurately code their results.

While the HL7 messaging standard does not currently address confidentiality or security issues, it is becoming widely accepted in the health care industry, and public health agencies are investigating its potential for transmitting surveillance data. Projects sponsored by CDC will use encryption, message authentication, and message nonrepudiation via a secure data facility to evaluate the ability to send and receive HL7 messages that meet or exceed current standards for the confidentiality and security of patient information.

The vocabulary selected for encoding laboratory information must also ensure that most reporting laboratories are able to participate and can accurately and efficiently code their results. This vocabulary will need to be unambiguous, expressive, comprehensive, and flexible in accommodating varying health care business practices and data structures. It should also eliminate redundancy, minimize maintenance, simplify mapping, and permit useful data aggregation and analysis.

LOINC has been frequently publicized and recommended by the HL7 community, yet it has not been widely implemented. The CDC is initiating several projects to evaluate the potential to use LOINC to encode test results data from laboratory information systems in various settings. To facilitate useful aggregation and analysis of the data, a second vocabulary or table is required to code non-numeric observation values such as organism names in OBX-5. The CDC will evaluate the potential to utilize SNOMED, which laboratories have used primarily to describe anatomic pathology data, to encode organism names and other non-numeric observation values in OBX-5.

Much of the information precoordinated in each LOINC code may be difficult or impossible to obtain within existing LISs and may be retrieved or inferred from other portions of the HL7 message. If the data structures or business rules of laboratories prevent them from implementing electronic laboratory reporting using LOINC as the observation identifier in OBX-3, alternative coding systems will be necessary. The HL7 standard could permit less fully specified observation identifiers in OBX-3, which in turn would permit LOINC to provide a table of precoordinated but simplified observation identifiers. While this would permit those facilities whose data structures or business rules cannot accommodate fully specified precoordinated observation identifiers to participate in public health electronic laboratory reporting, a means of coding non-numeric observation values would still be required. Another potential alternative is to use a postcoordinated vocabulary in OBX-3. This would allow the laboratory to compose as detailed a term in OBX-3 as their data structures permit without imposing the limitations of fully specified and precoordinated terms. Such simplification is OBX-3 could make the use of universal test identifiers genuinely feasible and reduce potential barriers to reporting electronic laboratory results to public health agencies. Some postcoordinated vocabularies such as SNOMED contain concepts that can also be used for observation values in OBX-5 and specimen codes in OBR-15.

Footnotes

This research was supported in part by an appointment to the Public Health Informatics Fellowship Program at the Centers for Disease Control and Prevention (CDC), administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and CDC.

[*] HL7 messages consist of variable-length data fields divided by a field separator character. The data fields are combined into logical groupings called "segments." The OBX segment contains information related to observations (such as laboratory test results). OBX-3 contains the observation (test) identifier.

▸ Top
▸ Abstract
▸ Introduction
▸ Public Health Reporting Characteristics
▸ Security Concerns
▸ Vocabulary Characteristics

References

1. Centers for Disease Control and Prevention. *Manual of Procedures for the Reporting of Nationally Notifiable Diseases to CDC*. Atlanta, GA: CDC, 1995.

2. Gostin L, Lazzarini Z, Neslund V, Osterholm M. The public health information infrastructure: a national review of the law on health information privacy. *JAMA*. 1996;24:1921-48.

3. Arden WF, McDonald CJ, Demoor G, et al. Logical observation identifier

- Implementation
- Conclusions
- References

names and codes (LOINC) database: a public use of set codes and names for electronic reporting of clinical laboratory test results: *Clin Chem*. 1996;42 (1):81-90. [PubMed]

4. Baorto DM, Cimino JJ, Parvin CA, Kahn MG. Using Logical Observation Identifier Names and Codes (LOINC) to exchange laboratory data among three academic hospitals: *Proc AMIA Annu Fall Symp*. 1997:96-100. [PubMed]
5. Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions and clarity. *J Am Med Inform Assoc*. 1997;4(3):238-51. [Free Full text in PMC]
6. Rothwell DJ, Côté RA, Cordeau JP, Boisvert MA. Developing a standard data structure for medical language: the SNOMED proposal. *Proc Annu Symp Comput Appl Med Care*. 1993:695-9. [PubMed]
7. Rothwell DJ, Côté RA. Managing information with SNOMED: understanding the model. *Proc AMIA Annu Fall Symp*. 1996:80-3. [PubMed]
8. Schulz EB, Price C, Brown PJ. Symbolic anatomic knowledge representation in the Read Codes version 3: structure and application. *J Am Med Inform Assoc*. 1997;4:38-48. [Free Full text in PMC]
9. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med*. 1998;37:394-403. [PubMed]
10. Cimino JJ. Formal descriptions and adaptive mechanisms for changes in controlled medical vocabularies. *Methods Inf Med*. 1996;35:202-10. [PubMed]
11. Cimino JJ, Clayton PD. Coping with changing controlled vocabularies. *Proc Annu Symp Comput Appl Med Care*. 1994:135-9. [PubMed]
12. Health Level Seven. *An Application Protocol for Electronic Data Exchange in Healthcare Environments*, version 2.3. Ann Arbor, Mich.: Health Level Seven, 1997.
13. Centers for Disease Conrol and Prevention. *Electronic Reporting of Laboratory Data for Public Health: Meeting Report and Recommendations*. Atlanta, GA: CDC, 1997. Available at: http://www.cdc.gov/phppo/dls/guidstd.htm. Accessed Mar 5, 1999
14. Centers for Disease Control and Prevention. *National Inventory of Clinical Laboratory Testing Services, Final Report*. Atlanta, Ga.: CDC, 1999. CDC contract 200-95-0933.
15. Aller R, Weilbert M, Carey K. Some LIS capabilities worth searching for. *CAP Today*. 1995;9(11):41-58. [PubMed]
16. Rocha RA, Huff SM. Coupling vocabularies and data structures: lessons from LOINC. *Proc AMIA Annu Fall Symp*. 1996:90-4. [PubMed]

Figures and Tables

Table 5. Examples of SNOMED Codes for Reporting in OBX-3

Write to PMC | PMC Home | PubMed
NCBI | U.S. National Library of Medicine
NIH | Department of Health and Human Services
Privacy Policy | Disclaimer | Freedom of Information Act

Methods of Information in Medicine
© F.K. Schattauer Verlagsgesellschaft mbH (2000)

Exhibit E

N. F. de Keizer, A. Abu-Hanna

Understanding Terminological Systems II: Experience with Conceptual and Formal Representation of Structure

Department of Medical Informatics,
Academic Medical Center, Amsterdam,
The Netherlands

Abstract: This article describes the application of two popular conceptual and formal representation formalisms, as part of a framework for understanding terminological systems. A precise understanding of the structure of a terminological system is essential to assess existing terminological systems, to recognize patterns in various systems and to build new terminological systems. Our experience with the application of this framework to five well-known terminological systems is described.

Keywords: Terminological System, Conceptual Representation, Formalization

1. Introduction

For many decades various terminological systems have been developed with different domains and different structures, such as strict hierarchies or semantic nets describing concepts and their relationships. These terminological systems were designed for different purposes. Cimino et al. [1] and Campbell et al. [2] describe criteria concerning the essential conceptual features of an ideal terminological system. Chute et al. recently summarized and extended these criteria [3]. In spite of papers giving an overview of the strengths and weaknesses of terminological systems [2, 4, 5] it is still hard to gain insight into the merits and usability of existing systems because the structure and characteristics of terminological systems are often incompletely and ambiguously described. We feel there is a need for a framework for *understanding* terminological systems, a framework which is still lacking. To understand and compare existing terminological systems and to evaluate them for specific goals there is a need for at least two components: (1) a uniform terminology and typology to characterize terminological systems themselves [6] and; (2) a uniform re-presentation formalism to describe the *structure* of these systems which is essential for understanding and evaluating existing terminological systems or for the development of new ones.

The goal of this paper is to provide a representation formalism for representing the structure of terminological systems and to report our experience with its application for formalizing existing terminological systems. Essential in our representation formalism is that it is conceptual, viz. it supports communication between, for example, domain experts and engineers of the terminological system. It should also help to highlight weak spots in the design by supporting the comparison of various terminological system structures and the comparison between characteristics of the systems with those characteristics required. Therefore, complementary to the conceptual part, a formal counterpart (based on first order logic) is needed to enhance expressivity and disambiguity and to support consistency during development of new terminological systems and their maintenance and reuse.

This article describes our experience with formalizing five well-known terminological systems: ICD [7, 8]; NHS clinical terms [9, 10]; SNOMED [11, 12]; UMLS [13, 14]; and GALEN [15, 16] and the comparison of these systems with criteria believed important for an ideal terminological system [1, 2].

In Section 2 we describe the representation formalism which is based on Entity Relationship Diagrams (ERD) and First Order Logic (FOL). In Section 3 the relevant criteria of Cimino et al. [1] and Campbell et al. [2] are translated into this formalism. In Section 4 the (simplified) structure of some well-known terminological systems are conceptually and formally described and these structures are compared with the criteria of an ideal terminological system. An extensive description can be found in a technical report [17] obtainable from the authors. In Section 5 we discuss similarities and differences among, and implications for the usability of these terminological systems.

2. Uniform Representation Formalism

A uniform representation formalism supports the complete and unambiguous description of the structure of a terminological system enabling comparison of different terminological systems described by the same formalism. Important characteristics of a uniform representation formalism are: (1) conceptuality – i.e. it lends itself for human comprehension and communication, (2) adequate expressive power, and (3) non-ambiguity.

The Entity-Relationship (ER) formalism [18] is a simple formalism capable of expressing concepts, relationships between concepts and some cardinality constraints, and it scores well on these criteria. The simple diagrammatic notations of ER (ER Diagram) have contributed immensely to its popularity. The notations use in the static part of object-oriented formalisms, such as OMT and UML, could have also been used, because they basically express the same content. However, ERD may not always be adequate for expressing complex constraints. Hence, a more expressive instrument and a formal (based on mathematical notions) specification, is needed to complement it in order to avoid non-ambiguity and to capture complex constraints. Based on its expressive power and universality we have chosen (many sorted) First-Order-Logic (FOL). Our choice for ER with FOL means that descriptions in this formalism could easily be translated to and from other logic-based formalisms such as Ontolingua [19], conceptual graphs [20] and description logics [21] when their expressivity allows this. Other researchers have described the use of logic-based formalisms for the representation of medical data and pointed out that this is a pre-condition for automated reasoning [20-24]. Therefore, a formal and conceptual representation of the terminological system's structure, that is a meta-model of the medical concepts it includes, is essential in understanding the system and hence its usability.

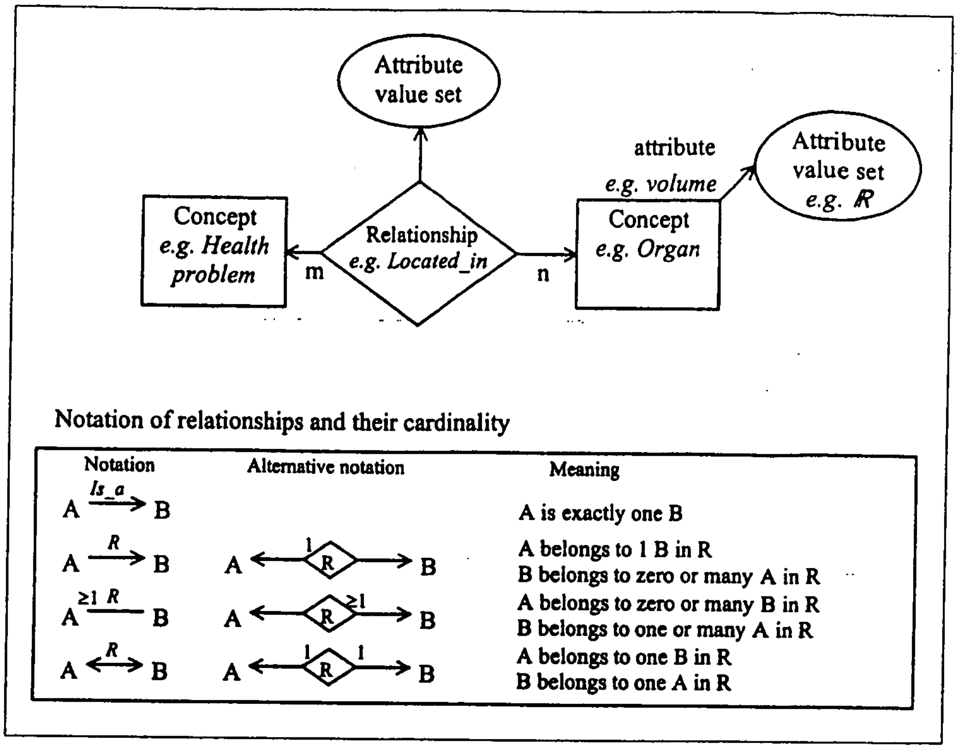


Fig. 1 ERD representation of concepts, attributes and relationships.

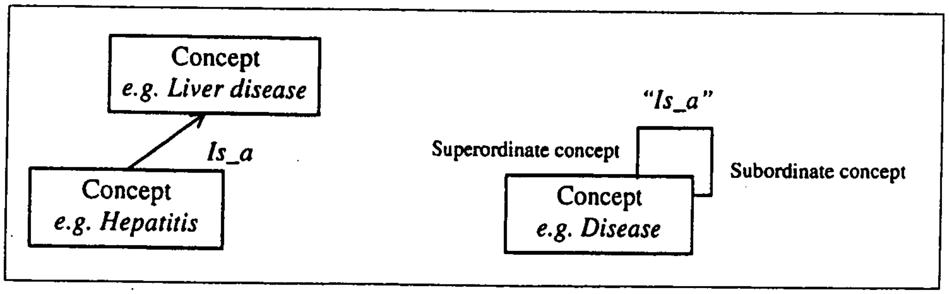


Fig. 2 ERD representation of hierarchical relationship normal (left) and at the meta-level (right).

3. Representation of Criteria for Terminological Systems

Two categories of criteria of an ideal terminological system [1, 2] can be distinguished: criteria which concern the representation formalism itself, and criteria which concern the descriptions of the domain (the model) using the representation formalism. In this section both categories are described with the ER formalism and a FOL description. An extensive description can be found in a previous technical report [17].

3.1 Criteria Concerning the Representation Formalism

In this section we describe criteria for the representation formalism which form the basis for the domain criteria described in Section 3.2. As explained in [6] the building blocks of most conceptualizations and hence also of terminological systems are *concepts*, *attributes* and *relationships* between concepts; these are represented in ERD by a rectangle, an arrow connected with an ellipse, and a diamond connecting rectangles by arrows, respectively (Fig. 1).

Relationships between concepts can be distinguished in hierarchical relationships ("Is_a" and "Is_part_of" relationships) and non-hierarchical relationships (e.g. "caused_by"). When modeling the medical domain we represent an "Is_a" relationship between two different concepts as shown on the left side of Fig. 2, e.g. hepatitis *Is_a* liver disease. For the purpose of the description of models of terminological systems themselves we distinguish in this paper a meta-level "Is_a" relationship (right side of Fig. 2). In the meta-model the meta-concept has instances which are concepts at the domain level, e.g. "hepatitis" and "liver disease" are hierarchically related concepts and are instances of the meta-concept "disease".

The rest of this section and Section 3.2 describe the criteria mentioned by Cimino et al. [1] and Campbell et al. [2]. A terminological system should enable the use of attributes to define or further specify concepts. *Relationships* between concepts should be *explicitly* represented by a label designating the meaning of the relationship, and constraints to restrict the interpretation of the relationship. Without explicitly representing relationships it is difficult to (automatically) interpret the meaning of a relationship, e.g., a relationship between "Health problem" and "Medication" can be interpreted as "Health problem *treated_by* Medication" or as "Health problem *is_side_effect_of* Medication". *Composition rules* are grammatical rules, which define how and which concepts can be used to compose new concepts. With FOL one could easily specify the composites declaratively.

3.2 Domain Model Criteria

This section describes the criteria which concern the model of terminological systems.

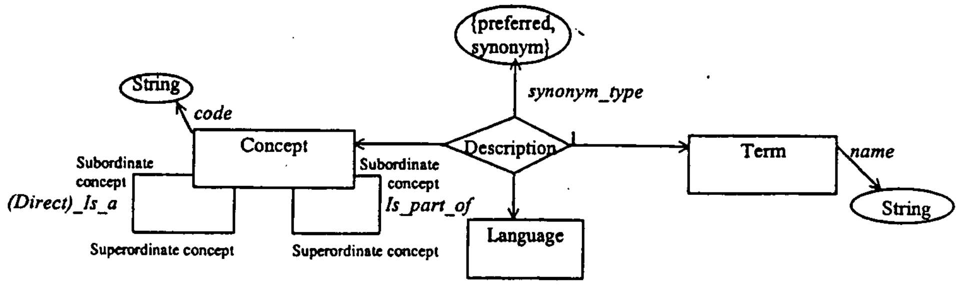
Domain completeness means that the terminological system should not be restricted in detail whether in depth or in breadth. Any constraint on the cardinality of the "Is_a" relationship or on the value of the (depth) level of the hierarchy, which reflects the chain of descendents of a concept, would hinder domain completeness (Fig. 3).

Synonyms and multi-lingual terms means that a unique concept may be designated by multiple terms in more than one language. Other model criteria concerning terms, concepts and their relationships are *non-ambiguity, non-vagueness* and *non-redundancy.* Non-vagueness prescribes that concepts must be complete in meaning, that is, refer to an object in the domain. Non-ambiguity prescribes that a concept refers to exactly one object in the domain. Non-redundancy prescribes that there should be a mechanism which prevents the existence of multiple different concept representations with the same meaning. These three criteria are not limited to the model of the terminological system, they are constraints on the meaning of concepts which ought to be considered by the knowledge engineer who develops the terminological system. To support *non-redundancy* in the model each concept has one preferred term and zero or more synonymous terms per language (see Fig. 3).

The next criterion *multiple classification* is represented by a hierarchical relationship in which a subordinate concept is related to as many superordinate concepts as required (see *Is_a* and *Is_part_of* relation in Fig. 3). For example "pneumococcal pneumonia" is a subordinate concept of the superordinate concept "lung diseases" as well as of the superordinate concept "infectious diseases".

A code can be conceived as an attribute of a concept. Codes must be non-significant, i.e. *context free* hence not related to the meaning or the position of the concept in the hierarchy. Another criterion concerning codes is the possibility of *cross-mapping,* e.g., for administrative reasons. This can be observed or realized by an attribute "cross-mapping code" appearing at each concept. The cross-mapping code e.g., between ICD and a local terminological system, can be either manually or (semi)automatically derived.

The last criterion concerns the *use of definitions.* Some terminological systems have textual definitions which have to be interpreted by the human reader. If these definitions were formal, a computer could (at least partially) process them and use them for automated reasoning such as consistency checking [21] and knowledge acquisition in GAMES [24] and PROTÉGÉ [25].



- $\forall_{c \in \text{Concept}} \forall_{l \in \text{Language}} \exists!_{t \in \text{Term}} (\text{Description}(c,l,t) \wedge \text{synonym\_type}(\langle c,l,t \rangle)=\text{preferred}))$
  For every language there is exactly one preferred term per concept. ∃!y A means that exactly one "y" exists for which A is true.
- $\forall_{c1 \in \text{Concept}} \forall_{l \in \text{Language}} \forall_{t \in \text{Term}} (\text{Description}(c1,l,t) \wedge \text{synonym\_type}(\langle c1,l,t \rangle)=\text{preferred}) \rightarrow$
  $\neg\exists_{c2 \in \text{Concept}} (\text{Description}(c2,l,t) \wedge \text{synonym\_type}(\langle c2,l,t \rangle)=\text{preferred} \wedge c2 \neq c1)$
  For each language a term can be preferred for just one concept, i.e. different concepts cannot have the same preferred term.

To avoid multiple classification one could specify:

- $\forall_{c1,c2,c3 \in \text{Concept}} (\text{Direct\_is\_a}(c1,c2) \wedge \text{Direct\_is\_a}(c1,c3) \rightarrow c3 = c2)$
  A subordinate concept c1 cannot have more than one superordinate concept

Fig. 3 ERD and FOL metadescription of non-redundancy, (multilingual) synonyms and multiple classification.

4. Description of Existing Terminological Systems

The conceptual and formal representation of terminological systems supports a better understanding of their structure. It helps to recognize the patterns in the designs of different terminological systems by a uniform view, which enables ascertainment of gaps and incompleteness in the terminological system.

In the first paper on this topic [6] we describe the typology and the coding scheme of the ICD-10 [8], NHS Clinical Terms [9, 10], SNOMED [11, 12], UMLS [13, 14] and GALEN [15, 16, 26]. We used ERD and FOL to describe the structure of these terminological systems and compare them with the criteria described in the previous section. For brevity, in this section we only represent the ERD and FOL description of ICD-10 and UMLS and only the essential information in FOL which is not represented in the ERD. An extensive conceptual and formal description of all systems is given in [17].

4.1 ICD-10

A conceptual and formal model of the ICD-10 is presented in Fig. 4.

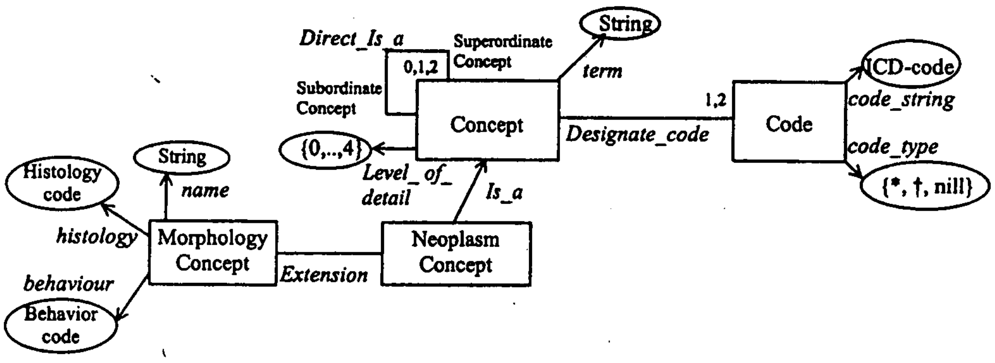
Although Fig. 4 shows explicit relationships to clarify the model of ICD-10, in reality the ICD-10 does not contain explicit relationships. For neoplasm concepts there are some term compositions. A coded nomenclature for morphology of neoplasms is part of the system. Each concept in the ICD-10 chapter 2 "Neoplasms" can be extended with a morphology concept and code, which consists of a histology code and a behavior code, e.g., "3=malignant, primary site". For non-neoplasm concepts there are no attributes and no composition rules to compose new complex concepts.

Domain completeness of ICD-10 is restricted to 4 levels of depth. Comparing Figures 3 and 4 shows that there is no real distinction between concepts and terms, hence synonyms, multilingual terms and non-redundancy are not supported. From the comparison between Figures 3 and 4 we also conclude that multiple classification is restricted in ICD-10 by the cardinality: concepts have 0, 1 or 2 parents. Each concept is at least designated by one unique code and at most by two unique codes: one dagger code related to the etiology and one asterisk code related to the location of the diagnostic term. Definitions, beyond the implicit interpretation that a subordinate concept is a more specific form of the superordinate concept, are lacking.

4.2 NHS Clinical Terms

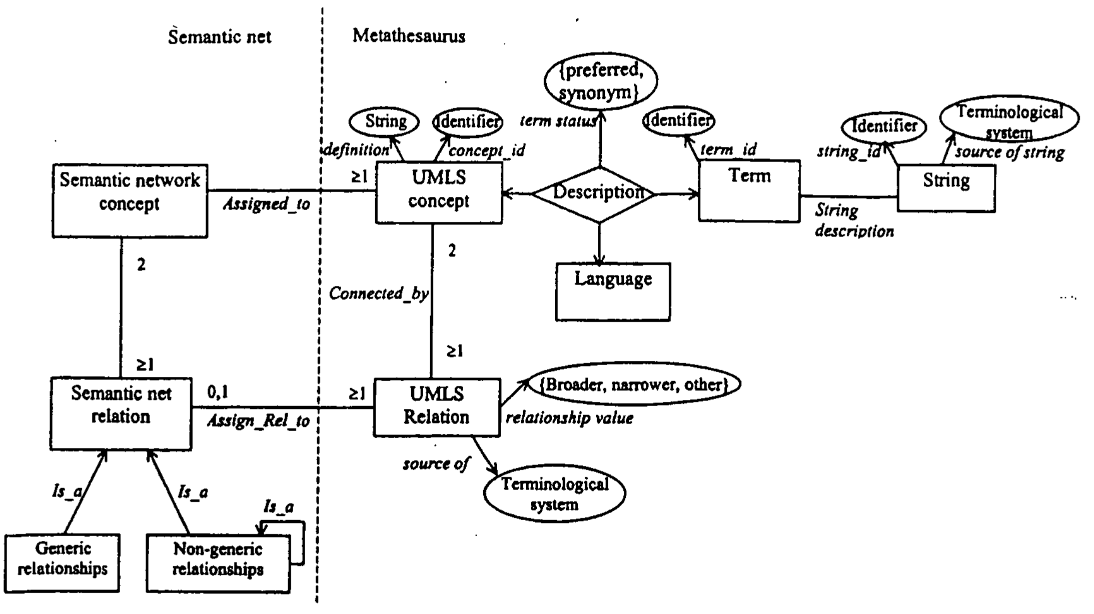
The NHS clinical terms, formerly known as the Read Clinical Classification forms a classification of medical concepts representing many domains such as diseases, signs, procedures, etc. Each of these subdomains contains concepts related by generic relationships. This system views partitive relationships as generic relationships by introducing *structure* concepts. For example, the subordinate concept "aortic arch" is part of the superordinate concept "thoracic aorta", but instead of a direct partitive link this concept is generically linked to the *structure* concept "thoracic aorta structure".

Although relationships are not formally made explicit, during the qualification of concepts, implicit relationships are used in the lookup tables (templates) to define combinations of concepts, attributes and attribute values in a controlled way [27]. The relationship between a concept and an attribute has a status: Qualifier if the attribute might supply extra detail which a user might choose to further describe a concept, e.g., "course of illness: (acute/chronic)" to qualify "heart failure"; Atom if the attribute is an intrinsic characteristic of a concept, e.g., "site: cardiac structure" of heart failure. In our terminology this is part of a defini-



- $\forall_{c \in Concept}$ Level_of_detail(c) ≤ 4
  The number of levels of Direct_Is_A (detail in depth) is restricted to four.
- $\forall_{c \in Concept} \forall_{cd \in Code}$ Designate_code(c,cd) ∧ | {<x,cd> ∈ Designate_code|x=c} | =1 → codetype(cd)=nill
  For each concept which is designated to exactly one code, that code has no specific type (nill)
- $\forall_{c \in concept} \forall_{cd1,cd2 \in code}$ Designate_code(c,cd1) ∧ Designate_code(c,cd2) ∧ cd1≠cd2 →
  ((code_type(cd1)= "*" ∧ code_type(cd2)= "†") ∨ (code_type(cd1)= "†" ∧ code_type(cd2)= "*"))
  If a concept is designated by two different codes, one of the codes has the type "*" and the other code has type "†"

Fig. 4 ERD and FOL representation of ICD-10.



- $\forall_{c\in \text{UMLSconcept}} \forall_{l\in \text{Language}} \exists!_{t\in \text{Term}}$ (Description(c,l,t) ∧ synonym_type(<c,l,t>)=preferred))
For every language there is exactly one preferred term per concept. ∃!y A means that exactly one "y" exists for which A is true.
- $\forall_{c1\in \text{UMLSconcept}} \forall_{l\in \text{Language}} \forall_{t\in \text{Term}}$ (Description(c1,l,t) ∧ synonym_type(<c1,l,t>)=preferred) →
$\neg\exists_{c2\in \text{Concept}}$ (Description(c2,l,t) ∧ synonym_type(<c2,l,t>)=preferred ∧ c2≠c1)
For each language a term can be preferred for just one concept, i.e. different concepts cannot have the same preferred term.
- $\forall_{c\in \text{UMLS concept}} \exists_{s\in \text{Semantic concept}}$ Assigned_to (c,s)
Each metathesaurus concept is assigned to at least one semantic concept in the semantic network

Fig. 5 ERD and FOL representation of the UMLS metathesaurus and semantic network.

tion. Template tables define which concepts can be modified by which attributes, but lack explicit constraints e.g., on combinations of attributes. Furthermore, existing atoms are not consistently used to describe the intension of their corresponding concept. Hence, definitions are unnecessarily partially specified.

NHS Clinical Terms has no limitations to domain completeness and multiple classification. Synonyms and unique codes are supported in the system. Each concept is designated by a description which consists of a unique code and a unique term (identifier) for the concept.

4.3 SNOMED

The structure of SNOMED consists of eleven modules (also called axes or dimensions), such as Topography, Disease and diagnosis, Procedures, etc., which can be conceived as distinct classifications. Concepts within one axis are related to each other using hierarchical relationships and concepts between different axes are also related by non-hierarchical relationships. An example of non-hierarchical relationships between terms are the relations between the disease "Tuberculosis" in the "Disease and diagnosis" and "Lung" in the topographical module, "Granuloma" in the morphology module, "M.tuberculosis" in the living organisms module and "Fever" in the function module. Since version II of SNOMED these relationships are explicit although a formal model lacks. Concepts are non-vague, that is they represent an object. Some terms, e.g., from the General modifier module, are vague but these are only used to compose new concepts. Concepts of the various modules/axes can be linked in order to compose new complex medical concepts, but SNOMED has not formalised any constraints on these compositions. SNOMED has no limitations to domain completeness and multiple classification. Multi-lingual terms, non-redundancy and unique codes are not supported. A disease concept can often be described by a non-vague concept from the "Disease and diagnoses" module, e.g., DE-14800 Tuberculosis, but in some cases it is also possible to describe the same concept with a concatenation of different concepts, e.g. Lung + Granuloma + M.tuberculosis + Fever which also represents Tuberculosis. Although each concept in SNOMED is described by one or more descriptions, there is no distinction between preferred and synonymous terms. Partial definitions are only available for concatenated concepts, in that case the definition is the enumeration of concatenated characteristics. SNOMED RT, which is under development, seems to be addressing all above-mentioned deficiencies by using a formal model [12] based on description logics [21].

4.4 UMLS

The UMLS consists of the Metathesaurus, the Semantic network, the SPECIALIST lexicon and the Information Sources Map. For brevity we only conceptually and formally describe the metathesaurus and the semantic network, because these two together form the most comparable component with the terminological systems described in the rest of this article. The Metathesaurus provides information about concepts, terms, string-names and the relationships between them, drawn from established terminological systems such as ICD-9-CM/ICD-10, SNOMED and MeSH. As shown in Figure 5 the metathesaurus represents "broader", "narrower", "other", relationships between different concepts (and optionally the relationships defined in the semantic network). Many relationships are derived directly from source terminological systems. For example, the hierarchical relationships in MeSH or ICD are manually made explicit in UMLS as "Is_a" or "Part_of" relationships. The UMLS does not contain composition rules to compose new complex concepts.

Comparing Figures 3 and 5 shows that UMLS has no limitations on domain completeness, and that multiple classification is possible. It also shows that synonyms, multiple-lingual terms, non-redundancy and unique codes are supported in UMLS. The formal specification of Figure 5 shows that each concept in the UMLS is described by one preferred and possibly more synonymous terms which are in turn linked to multiple strings (plurals, etc.). Each concept has an attribute "definition", a textual definition which describes the meaning of the concept.

Each concept in the metathesaurus is assigned to the most (one of 132) specific semantic category available in the semantic network. The semantic network provides information about the set of basic semantic categories (also called semantic types): Physical objects (e.g., organisms), Conceptual entities (e.g., findings), Activities (e.g., behavior) and Phenomenons and processes (e.g., biological function) and their relationships. Via the "Is_a" link, relationships and attributes are inherited by the subordinates of the high level semantic category. By inheritance the relationship "process_of" between "Biological function" and "Organism" also exists between "Disease or Syndrome" (which *Is_a* "Biologic Function") and "Human" (which *Is_a* an "Organism"). The inheritance of relationships can be blocked in case the subordinates of a semantic category conflict with the relationship, e.g. "Mental or Behavioral dysfunction" is a "Biologic Function" which can be related to an "Organism" by the "process of" relationship. "Plants" is a subordinate of "Organisms" but cannot have a mental dysfunction, therefore this inheritance is blocked. Relationships between semantic categories do not necessarily apply to all metathesaurus concepts that have been assigned to those semantic categories. For example, the relationship *"evaluation_of"* exists between the semantic categories "Sign" and "Organism attribute". The metathesaurus concept "overweight", related to the semantic category "Sign", is an *evaluation of* the "Organism attribute" concept "body weight" but it is not an *evaluation of* the "Organism attribute" concept "body length". Inconsistencies between metathesaurus concepts cannot be blocked.

4.5 GALEN

GALEN is different from the abovementioned terminological systems. Like UMLS and SNOMED RT, GALEN provides an explicit model of the domain but it also provides a flexible representation language. GALEN's goal is to formally describe and model the medical domain by which the interchangeability of electronic medical data of different data sources can be supported. The Concept Module utilizes the GALEN Representation and Integration Language (GRAIL), a formalism based on Description Logics, to represent and manipulate the Concept Reference (CORE) model. The Core model is an ontology currently comprising approximately 5,000 concepts and 1,000 explicit relationships. The GRAIL formalism allows developers of terminologies to create models containing these concepts and relationship, and to derive (automatically) new composed concepts. There are no restrictions on domain completeness. To guarantee non-redundant and sensible composed concepts, automated reasoning facilitated by reasoning services of Description Logics [21], is used. Relationships be-tween concepts can be: "sanctioning", which specify how concepts are allowed to be used in the formation of composites; or "descriptive", which specify intrinsic characteristics of a concept. A conceptual and formal representation of a general GRAIL model is described in the technical report related to this article [17]. Automated reasoning within GRAIL facilitates multiple classification. Models developed with GRAIL are language independent and therefore the model of concepts is separated from the (synonymous) terms used to designate them. The Multilin-gual Module manages the mapping of concepts to preferred and synonymous terms of different languages. The Code Conversion Module can be used as an inter-lingua and manages the mapping of concepts to and from existing coding systems.

5. Discussion

In this article we have described the second part of a framework for understanding terminological systems and our experiences with its application: the use of conceptual and formal representation formalisms for representing the structure of a terminological system. After applying the formalism to the required criteria of an ideal terminological system (Section 3), we applied the representation formalism to describe the structure of important terminological systems (Section 4). In our experience, formalization supports the comparison between the criteria on terminological systems (Figs. 1 to 3) and the structure of existing terminological systems (Figs. 4 and 5). Formalization resulted in a reference design that helped us to observe discrepancies in some terminological systems. The largest problems with the ICD, which we observed during this exercise, are the lack of explicit relationships and definitions, the lack of separating terms and concepts (and so the lack of synonyms), the lack of possibilities for composing new concepts and the restrictions of domain completeness (number of levels in depth restricted to 4). We think that especially (formal) definitions and composition rules are essential criteria for future terminological systems because, especially when expressed in a restricted form, they can facilitate automated reasoning such as consistency checking, classification [21] and knowledge acquisition (data entry) [24, 25].

The NHS Clinical Terms, SNOMED and the UMLS do better on most criteria, but composition rules and formal definitions are also missing or premature. SNOMED RT aims to address this deficiency but it is not yet operational. The UMLS uses a semantic network to structure concepts in the medical domain. Reasoning is not (optimally) supported in the UMLS because nodes and arcs are only intuitively defined by their labels on a high level of detail, and therefore sensibility control on metathesaurus concept level is not supported.

The GALEN project, not extensively operational yet, is an ambitious and promising project aiming at a formal description of the total medical domain. GALEN has the intention to adhere to the criteria mentioned in Section 3 but the realization of all these intentions is still under construction. Moreover, the restrictions of the description logics on which GALEN is based, improves reasoning with the system, but constrains its expressivity. But even the somewhat restricted expressiveness of GALEN and its orientation towards the total medical domain still implies that a great effort should go into the syntax and grammatical rules to guarantee sensibly composed concepts. We intend a thorough and practical evaluation of the CORE model and the GALEN applications in a collaborative research. Special attention will be paid to the evaluation of the expressiveness and the possibilities for automated reasoning in clinical settings.

We believe that the formalization of terminological systems helped us with a thorough understanding of their structure and merits. It was also the basis for the development of a new terminological system for intensive care [28]. In our opinion, formal representation formalisms constitute indispensable tools for serious terminological system developers. The representation of the structure of a terminological system in a conceptual and formal way has more advantages next to merely *understanding* the terminological system [20, 22-24, 29]. A conceptual and formal representation of a structure of a system supports the communication about what the system means and it supports development of new systems by finding the "desired patterns" (criteria of Section 3) in the design and by building new designs based on desired ones. Furthermore, by making knowledge of the underlying domain explicit with formal specifications, these specifications can be used in a knowledge acquisition tool, such as GAMES [24] and PROTÉGÉ [25], to support inference of new knowledge and to support consistency checks within the terminological system. These are important tools for the management of the terminological system.

Acknowledgement

We thank Ronald Cornet for his valuable discussions and helpful remarks on an earlier draft of this manuscript.

REFERENCES

1. Cimino J, Clayton P, Hripcsak G, Johnson S. Knowledge-based approaches to the maintenance of a large controlled medical terminology. J Am Med Informatics Assoc 1994; 1: 35-50.
2. Campbell J, Carpenter P, Sneiderman C, et al. Phase II evaluation of clinical coding schemes: completeness, definitions and clarity. J Am Med Inform Assoc 1997; 4: 238-51.
3. Chute C, Cohn S, Campbell J, et al. A framework for comprehensive Health Terminology Systems in the United States: Development guidelines, criteria for selection and public policy implications. J Am Med Inform Assoc 1998; 5 (6): 503-10.
4. Cimino J. Coding Systems in Health Care. Yearbook of Medical Informatics 1995: 71-85.
5. Klimczak J, Hahn A, Sievert M, Petroski G, Hewett J. Comparing clinical vocabularies using coding system fidelity. In: Gardner R, ed. Annual Symposium on Computer Applications in Medical Care. New Orleans: Hanley & Belfus 1995.
6. de Keizer N, Abu-Hanna A, Zwetsloot-Schonk J. Understanding terminological systems I: terminology and typology. Method Inform Med 2000; 39: 16-21.
7. International Classification of Diseases, manual of the International Statistical Classification of diseases, injuries and causes of death: 9th revision. WHO 1977.
8. International Classification of Diseases, manual of the International Statistical Classification of diseases, injuries and causes of death: 10th revision. WHO 1993.
9. Robinson D, Comp D, Schulz E, Brown P, Price C. Updating the Read Codes: User-interactive Maintenance of a Dynamic Clinical Vocabulary. J Am Med Inform Assoc 1997; 4 (6): 465-72.
10. Schulz E, Price C, Brown P. Symbolic Anatomic Knowledge Representation in the Read Codes Version 3: Structure and Application. J Am Med Inform Assoc 1997; 4: 38-48.
11. Rothwell D. SNOMED-Based knowledge representation. Method Inform Med 1995; 34: 209-13.
12. Spackman K, Campbell K. Compositional concept representation using SNOMED: towards further convergence of clinical terminologies. Am Med Inform Assoc annual symposium, 1998: 740-4.
13. UMLS Knowledge sources – Documentation. National Library of Medicine 1997:
14. Lindberg D, Humphreys B, Mc Cray A. The Unified Medical Language System. Method Inform Med 1993; 34: 281-91.
15. Rector A, Solomon W, Nowlan W, Rush T, Zanstra P, Claassen W. A Terminology Server for medical language and medical information systems. Method Inform Med 1995; 34: 147-57.
16. Rector A, Bechhofer S, Goble C, Horrocks I, Nowlan W, Solomon W. The Grail concept modelling language for medical terminology. Artif Intell 1997; 9: 139-71.
17. de Keizer N, Abu-Hanna A. A framework for understanding Terminological Systems. Amsterdam: Academic Medical Center. Dep Medical Informatics 1999.
18. Chen P. The Entity-Relationship model; toward a unified view of data. ACM Trans on Database Systems 1966; n: vol 1:1.
19. Gruber T. A translation approach to portable ontology specifications. Knowledge Acquisition 1993; 5: 199-220.
20. Campbell K, Das A, Musen M. A logical foundation for representation of clinical data. J Am Med Informatics Assoc 1994; 1: 218-32.
21. Donini F, Lenzerini M, Nardi D et al. Reasoning in Description Logics. In: Brewka G, ed. Principles of knowledge representation and reasoning. CLSI publications, 1996: 193-238.
22. Abu-Hanna A, Jansweijer W. Modeling application domain knowledge using explicit conceptualization. IEEE-Expert 1994.
23. Gruber T. Towards principles for the design of ontologies used for knowledge sharing. Int J Hum Comp Stud 1995; 43: 907-28.
24. Heijst G, Falasconi S, Abu-Hanna A, Schreiber A, Stefanelli M. A case study in ontology library construction. Art Intell Med 1995; 7: 227-55.
25. Studer R, Eriksson H, Gennari J, Tu S, Fensel D, Musen M. Ontologies and the Configuration of Problem-solving Methods. In: Gains B, Musen M, eds. 10th Banff Knowledge Acquisition for Knowledge-

Based Systems Workshop. Banff, Canada 1996.

26. Rector A, Glowinski A, Nowlan W, et al. Medical-concept models and medical records: an approach based on GALEN and PEN&PAD. J Am Med Inform Assoc 1995; 2: 19-35.
27. Read Code File structure version 3.1. NHS Centre for Coding and Classification 1995.
28. Campbell K, Oliver D, Spackman K, Shortliffe K. Representing Thoughts, Words, and Things in the UMLS. J Am Med Inform Assoc 1998; 5 (5): 421-31.
29. Moorman P, van Ginneken A, van der Lei J, van Bemmel JH. A model for structured data entry based on explicit descriptional knowledge. Yearbook of Medical Informatics 1995: 195-204.

Address of the authors:
Nicolette F. de Keizer,
Department of Medical Informatics, J2-256,
Academic Medical Center, P.O. Box 22660,
1100 DD Amsterdam, The Netherlands
E-mail: n.f.keizer@amc.uva.nl

Methods of Information in Medicine

Exhibit F

J. J. Cimino

Desiderata for Controlled Medical Vocabularies in the Twenty-First Century

Department of Medical Informatics,
Columbia University, New York, USA

Abstract: Builders of medical informatics applications need controlled medical vocabularies to support their applications and it is to their advantage to use available standards. In order to do so, however, these standards need to address the requirements of their intended users. Over the past decade, medical informatics researchers have begun to articulate some of these requirements. This paper brings together some of the common themes which have been described, including: vocabulary content, concept orientation, concept permanence, nonsemantic concept identifiers, polyhierarchy, formal definitions, rejection of "not elsewhere classified" terms, multiple granularities, multiple consistent views, context representation, graceful evolution, and recognized redundancy. Standards developers are beginning to recognize and address these desiderata and adapt their offerings to meet them.

Keywords: Controlled Medical Terminology, Vocabulary, Standards, Review

1. Introduction

The need for controlled vocabularies in medical computing systems is widely recognized. Even systems which deal with narrative text and images provide enhanced capabilities through coding of their data with controlled vocabularies. Over the past four decades, system developers have dealt with this need by creating ad hoc sets of controlled terms for use in their applications. When the sets were small, their creation was a simple matter, but as applications have grown in function and complexity, the effort needed to create and maintain the controlled vocabularies became substantial. With each new system, new efforts were required, because previous vocabularies were deemed unsuitable for adoption in or adaptation to new applications. Furthermore, information in one system could not be recognized by other systems, hindering the ability to integrate component applications into larger systems.

Consider, for example, how a computer-based medical record system might work with a diagnostic expert system to improve patient care. In order to achieve optimal integration of the two, transfer of patient information from the record to the expert would need to be automated. In one attempt to do so, the differences between the controlled vocabularies of the two systems was found to be the major obstacle – even when both systems were created by the same developers [1].

The solution seems obvious: standards [2]. In fact, many standards have been proposed, but their adoption has been slow. Why? System developers generally indicate that, while they would like to make use of standards, they can't find one that meets their needs. What are those needs? The answers to this question are less clear. The simple answer is, "It doesn't have what I want to say." Standards developers have taken this to mean that the solution is equally simple: keep adding terms to the vocabulary until it does say what's needed. However, systems developers, as users of controlled vocabularies, are like users everywhere: they may not always articulate their true needs. Vocabulary developers have labored to increase their offerings, but have continued to be confronted with ambivalence. A number of vocabularies have been put forth as standards [3] but they have been found wanting in some recent evaluations [4-6].

Over the past ten years or so, medical informatics researchers have been studying controlled vocabulary issues directly. They have examined the structure and content of existing vocabularies to determine why they seem unsuitable for particular needs, and they have proposed solutions. In some cases, proposed solutions have been carried forward into practice and new experience has been gained. As we prepare to enter the twenty-first century, it seems appropriate to pause to reflect on this additional experience, to rethink the directions we should pursue, and to identify the next set of goals for the development of standard, reusable, multipurpose controlled medical vocabularies.

2. Desiderata

The task of enumeration of general desiderata for controlled vocabularies is hampered in two ways. First, the desired characteristics of a vocabulary will vary with the intended purpose of that vocabulary and there are many possible intended purposes. I address this issue by stating that the desired vocabulary must be multipurpose. Some of the obvious purposes include: capturing clinical findings, natural language processing, indexing medical records, indexing medical literature, and representing medical knowledge. Each reader can add his or her own favorite purpose. A vocabulary intended for any of these can, and often has been, created. But the demands placed on a vocabulary become very different when it must meet several purposes.

A second obstacle to summarizing general desiderata is the difficulty teasing out individual opinions from the literature and unifying them. The need for controlled vocabularies for medical computing is almost as old as computing itself [7-9]. However, it is only in the past ten years or so that researchers have gone past talking about the content of a vocabulary and started to talk about deeper representational aspects. Before then, the literature contains many implications and half-stated assertions. Since then, authors have become more explicit, but the "terminology of terminology" has not yet settled down to a level of general understanding about what each of us means when we discuss vocabulary characteristics [10]. It is a foregone conclusion, then, that the summary I present here is bound to misrepresent some opinions and overlook others.

With these disclaimers, then, I will attempt to enumerate some of the characteristics that seem to be emerging from recent vocabulary research. Some of them may seem obvious, but they are listed formally in order that they not be overlooked.

2.1 Content, Content, and Content

Like the three most important factors in assessing the value of real estate (location, location, and location), the importance of vocabulary content can not be over stressed. The first criticisms of vocabularies were almost universally for more content. The need for expanded term coverage continues to be a problem, as can be seen in numerous studies which evaluate available standards for coverage of a particular domains. For example, recent publications examining the domain of nursing terminology are almost completely focused on the issue of what can be said [11-13]. Issues such as how things can be said, or how the vocabulary is organized are apparently less urgent, although sometimes solutions may need to go beyond the simple addition of more terms [14, 15].

One approach to increasing content is to add terms as they are encountered, responding as quickly as possible to needs as they arise [16]. In this approach, one adds complex expressions as needed rather than attempting a systematic, anticipatory solution. For example, rather than try to anticipate every kind of fracture (simple vs. compound, greenstick vs. avulsion vs. compression, etc.) of every bone, one would add terms for the most common and add more as needed. This avoids the large numbers of terms occurring through combinatorial explosion and the enumeration of nonsensical combinations (such as "compound greenstick fracture of the stapes", an anatomically implausible occurrence for a small bone in the middle ear).

An alternative approach is to enumerate all the atoms of a terminology and allow users to combine them into necessary coded terms [17], allowing compositional extensibility [18]. The tradeoff is that, while domain coverage may become easier to achieve, use of the vocabulary becomes more complex. Even with this atomic approach, the identification of all the atoms is nontrivial. The atoms must be substantial enough to convey intended meaning and to preserve their meaning when combined with others. They must be more than simply the words used in medicine. For example, the atom "White" could be used for creating terms like "White Conjunctiva" but would be inappropriate to use in constructing terms such as "Wolff-Parkinson-White Syndrome". The word "White" needs to be more than a collection of letters – after all, we could represent all medical concepts with just the letters of the alphabet (26, more or less), but this would hardly advance the field of medical informatics. The atomic approach must also consider how differentiate between atoms and mo cules. "White" and "Conjunctiva" a almost certainly atoms, but what abc "Wolff-Parkinson-White Syndrome"

No matter what approach is take the need for adding content remai This occurs because users will dema additions as usage expands and becau the field of medicine (with its attenda terminology) expands. The real issue address in considering the "conte desideratum" is this: a formal meth dology is needed for expanding conter A haphazard, onesy-twosy approa usually fails to keep up with the nee of users and is difficult to apply cons stently. The result can be a patchwo of terms with inconsistent granulari and organization. Instead, we nee formal, explicit, reproducible metho for recognizing and filling gaps i content. For example, Musen et a applied a systematic approach (negotia tion of goals, anticipation of use, accom modation of a user community, an evaluation) to creation of a vocabular for use in a progress note system [19] Methods of similar rigor need to b developed which can be used fo content discovery and expansion i large, multipurpose vocabularies. Mor attention must be focused on how representations are developed, rathe than what representations are produce [20].

2.2 Concept Orientation

Careful reading of medical informatics research will show that most systems that report using controlled vocabulary are actually dealing with the notion of concepts. Authors are becoming more explicit now in stating that they need vocabularies in which the unit of symbolic processing is the concept – an embodiment of a particular meaning [21-25]. Concept orientation means that terms must correspond to at least one meaning ("nonvagueness") and no more than one meaning ("nonambiguity"), and that meanings correspond to no more than one term ("nonredundancy") [26, 27].

Review of the literature suggests that there is some argument around the issue of ambiguity. Blois argues that while low-level concepts (such as protons) may be precisely defined, high-level concepts, including clinical concepts like myocardial infarction, are defined not by necessary attributes but by contingent ones (e.g., the presence of chest pain in myocardial infarction) [28]. Moorman et al. suggest that ambiguity can be allowed in the vocabulary as long as it is reduced to unequivocal meaning, based on context, when actually used (e.g., stored in a clinical record) [29].

However, a distinction must be made between ambiguity of the meaning of a concept and ambiguity of its usage [26, 30]. It is unfair, for example, to say that the concept "Myocardial Infarction" is ambiguous because it could mean "Right Ventricular Infarction", "Left Ventricular Infarction" and so on. Any concept, no matter how fine-grained, will always subsume some finer-grained concepts. But "Myocardial Infarction" has a meaning which can be expressed in terms of a particular pathophysiologic process which affects a particular anatomic site. Now, if we use this concept to encode patient data, the meaning of the data will vary with the context ("Myocardial Infarction", "Rule Out Myocardial Infarction", "History of Myocardial Infarction", "Family History of Myocardial Infarction", "No Myocardial Infarction", etc.).

This context-sensitive ambiguity is a different phenomenon from context-independent ambiguity that might be found in a controlled vocabulary [31]. For example, the term "Diabetes" does not subsume "Diabetes Mellitus" and "Diabetes Insipidus"; it has no useful medical meaning (vague). The concept "MI" might mean "Myocardial Infarction", "Mitral Insufficiency", or "Medical Informatics"; before it even appears in a context, it has multiple meanings (ambiguous). Concept orientation, therefore, dictates that each concept in the vocabulary has a single, coherent meaning, although its meaning might vary, depending on its appearance in a context (such as a medical record) [29].

2.3 Concept Permanence

The corollary of concept orientation is concept permanence: the meaning of a concept, once created, is inviolate. Its preferred name may evolve, and it may be flagged inactive or archaic, but its meaning must remain. This is important, for example, when data coded under an older version of the vocabulary need to be interpreted in view of a current conceptual framework. For example, the old concept "pacemaker" can be renamed "implantable pacemaker" without changing its meaning (as we add the concept "percutaneous pacemaker"). But, the name for the old concept "non-A-non-B hepatitis" can not be changed to "Hepatitis C" because the two concepts are not exactly synonymous (that is, we can't infer that someone diagnosed in 1980 as having non-A-non-B hepatitis actually had hepatitis C). Nor can we delete the old concept, even though we might no longer code patient data with it.

2.4 Nonsemantic Concept Identifier

If each term in the vocabulary is to be associated with a concept, the concept must have a unique identifier. The simplest approach is to give each concept a unique name and use this for the identifier. Now that computer storage costs are dropping, the need for the compactness provided by a code (such as an integer) has become less compelling. If a concept may have several different names, one could be chosen as the preferred name and the remainder included as synonyms. However, using a name as a unique identifier for a concept limits our ability to alter the preferred name when necessary. Such changes can occur for a number of reasons without implying that the associated meaning of the concept has changed [32].

Because many vocabularies are organized into strict hierarchies, there has been an irresistible temptation to make the unique identifier a hierarchical code which reflects the concept's position in the hierarchy. For example, a concept with the code 1000 might be the parent of the concept with the code 1200 which, in turn might be the parent of the concept 1280, and so on. One advantage to this approach is that, with some familiarity, the codes become somewhat readable to a human and their hierarchical relationships can be understood. With today's computer interfaces, however, there is little reaso why humans need to have readab codes or, for that matter, why they eve need to see the codes at all. Anothe advantage of hierarchical codes is tha querying a database for members o a class becomes easier (e.g., searchin for "all codes beginning with 1" wi retrieve codes 1000, 1200, 1280, an so on). However, this advantage is los if the concepts can appear in multipl places in the hierarchy (see "Polyhierarchy", below); fortunately, there ar other ways to perform "class-based" queries to a database which will wor even when concepts can be in multipl classes [32].

There are several problems wit using the concept identifier to conve hierarchical information. First, it is possible for the coding system to run out of room. A decimal code, such as the one described above, will only allow ten concepts at any level in the hierarchy and only allow a depth of four [34]. Coding systems can be designed to avoid this problem, but other problems remain. For example, once assigned a code, a concept can never be reclassified without breaking the hierarchical coding scheme. Even more problematic, if a concept belongs in more than one location in the hierarchy (see "Polyhierarchy", below), a convenient single hierarchical identifier is no longer possible. It is desirable, therefore, to have the unique identifiers for the concepts which are free of hierarchical or other implicit meaning (i.e., nonsemantic concept identifiers); such information should instead be included as attributes of the concepts [14].

2.5 Polyhierarchy

There seems to be almost universal agreement that controlled medical vocabularies should have hierarchical arrangements. This is helpful for locating concepts (through "tree walking"), grouping similar concepts, and conveying meaning (for example, if we see the concept "cell" under the concept "anatomic entity" we will understand the intended meaning as different than if it appeared under the concepts "room" or "power source"). There is some disagreement, however, as to whether concepts should be classified according to a single taxonomy (strict hierarchy) or if multiple classifications (polyhierarchy) can be allowed. Most available standard vocabularies are strict hierarchies.

Different vocabulary users may demand different, equally valid, arrangements of concepts. It seems unlikely that there can ever be agreement on a single arrangement that will satisfy all; hence the popular demand for multiple hierarchies [31, 34-36]. Zweigenbaum and his colleagues believe that concept classification should be based on the essence of the concepts, rather than arbitrary descriptive knowledge [37]. They argue quite rightly that arbitrary, user-specific ad hoc classes can still be available using additional semantic information. However, unless there can be agreement on what the essence of concepts should be, there can never be agreement on what the appropriate hierarchy should be. Furthermore, if the essence of a concept is defined by its being the union of the essence of two other concepts, its classification becomes problematic. For example, until medical knowledge advances to provide a better definition, we must define the essence of "hepatorenal syndrome" as the occurrence of renal failure in patients with severe liver disease. If our vocabulary has the concepts "liver disease" and "renal disease" (which seem desirable or at least not unreasonable), "hepatorenal syndrome" must be a descendant of both.

There can be little argument that strict hierarchies are more manageable and manipulable, from a computing standpoint, than polyhierarchies. This is small consolation, however, if the vocabulary is unusable. General consensus, seems to favor allowing multiple hierarchies to coexist in a vocabulary without arguing about which particular tree is the essential one. It is certainly possible that if a single hierarchy is needed for computational purposes, one could be so designated with the others treated as nonhierarchical (but nevertheless directed and acyclic) relationships.

2.6 Formal Definitions

Many researchers and developers have indicated a desire for controlled vocabularies to have formal definitions in one form or another [23, 25-27, 36, 38-50]. Usually, these definitions are expressed as some collection of relationships to other concepts in the vocabulary. For example, the concept "Pneumococcal Pneumonia" can be defined with a hierarchical ("is a") link to the concept "Pneumonia" and a "caused by" link to the concept "Streptococcus pneumoniae". If "Pneumonia" has a "site" relationship with the concept "Lung", then "Pneumococcal Pneumonia" will inherit this relationship as well. This information can be expressed in a number of ways, including frame-based semantic networks [40], classification operators [51], categorical structures [52], and conceptual graphs [53-55]. The important thing to realize about these definitions is that they are in a form which can be manipulated symbolically (i.e., with a computer), as opposed to the unstructured narrative text variety, such as those found in a dictionary. Many researchers have included in their requests that the definitional knowledge be made explic-itly separated from assertional knowledge which may also appear in the vocabulary [25, 41, 43, 46, 56]. For example, linking "Pneumococcal Pneumonia", via the "caused by" relationship, to "Streptococcus pneumoniae" is definitional, while linking it, via a "treated with" relationship, to "Penicillin" would be assertional. Similarly, the inverse relationship ("causes") from "Streptococcus pneumoniae" to "Pneumococcal Pneumonia" would also be considered assertional, since it is not part of the definition of "Streptococcus pneumoniae".

The creation of definitions places additional demands on the creators of controlled vocabularies. However, with careful planning and design, these demands need not be onerous. For example, the definition given for "Pneumococcal Pneumonia", given above, only required one additional "caused by" link to be added, assuming that it would be made a child of "Pneumonia" in any case and that the concept "Streptococcus pneumoniae" was already included in the vocabulary. Many of the required links can be generated through automatic means, either by the processing of the concept names directly [18] or through extraction from medical knowledge bases [57]. Also, the effort required to include definitions may help not only the users of the vocabulary, but the maintainers as well: formal definitions can support automated vocabulary management [58], collaborative vocabulary development [59], and methods for converging distributed development efforts [60, 61].

2.7 Reject "Not Elsewhere Classified"

Since no vocabulary can guarantee domain completeness all of the time, it is tempting to include a catch-all term which can be used to encode information that is not represented by other existing terms. Such terms often appear in vocabularies with the phrase "Not Elsewhere Classified", or "NEC" (this is not to be confused with "Not Otherwise Specified", or "NOS", which simply means that no modifiers are included with the base concept). The problem with such terms is that they can never have a formal definition other than one of exclusion – that is, the definition can only be based on knowledge of the rest of concepts in the vocabulary. Not only is this awkward, but as the vocabulary evolves, the meaning of NEC concepts will change in subtle ways. Such "semantic drift" will lead to problems, such as the proper interpretation of historical data. Controlled vocabularies should therefore reject the use of "not elsewhere classified" terms.

2.8 Multiple Granularities

Each author who expresses a need for a controlled vocabulary, does so with a particular purpose in mind. Associated with that purpose, usually implicitly, is some preconception of a level of granularity at which the concepts must be expressed. For example, the concepts associated with a diabetic patient might be (with increasingly finer granularity): "Diabetes Mellitus", "Type II Diabetes Mellitus", and "Insulin-Dependent Type II Diabetes Mellitus" (note that the simpler term "Diabetes" is so coarse-grained as to be vague). A general practitioner might balk at being required to select a diagnosis from the fine-grained end of this spectrum of concepts, while an endocrinologist might demand nothing less.

In reviewing the various writings on the subject, it becomes clear that multiple granularities are needed for multipurpose vocabularies. Vocabularies which attempt to operate at one level of granularity will be deemed inadequate for application where finer grain is needed and will be deemed cumbersome where coarse grain is needed. Insistence on a single level of detail within vocabularies may explain why they often are not reusable [62]. It also conflicts with a very basic attribute of medical information: the more macroscopic the level of discourse, the coarser the granularity of the concepts [63].

It is essential that medical vocabularies be capable of handling concepts as fine-grained as "insulin molecule" and as general as "insulin resistance". However, we must differentiate between the precision in medical knowledge and the precision in creating controlled concepts to represent that knowledge. While uncertainty in medical language is inevitable [64], we must strive to represent that uncertainty with precision.

2.9 Multiple Consistent Views

If a vocabulary is intended to serve multiple functions, each requiring a different level of granularity, there will be a need for providing multiple views of the vocabulary, suitable for different purposes [30]. For example, if an application restricts coding of patient diagnoses to coarse-grained concepts (such as "Diabetes Mellitus"), the more fine-grained concepts (such as "Insulin-Dependent Type II Diabetes Mellitus") could be collapsed into the coarse concept and appear in this view as synonyms (see Figs. 1a and 1b). Alternatively, an application may wish to hide some intermediate classes in a hierarchy if they are deemed irrelevant (see Fig. 1c). Similarly, although the vocabulary may support multiple hierarchies, a particular application may wish to limit the user to a single, strict hierarchy (see Fig. 1d).

We must be careful to confine the ability to provide multiple consistent views, such that inconsistent views do not result. For example, if we create a view in which concepts with multiple parents appear in several places in a single hierarchy, care must be taken that each concept has an identical appearance within the view (see Figs. 1e and 1f) [31].

2.10 Beyond Medical Concepts: Representing Context

Part of the difficulty with using a standard controlled vocabulary is that the vocabulary was created independent of the specific contexts in which it is to be used. This helps prevent the vocabulary from including too many implicit assumptions about the meanings of concepts and allows it to stand on its own. However, it can lead to confusion when concepts are to be recorded in some specific context, for example, in an electronic patient record. Many researchers have expressed a need for their controlled vocabulary to contain context representation through formal, explicit information about how concepts are used [21, 65, 66].

A decade ago, Huff and colleagues argued that a vocabulary could never be truly flexible, extensible and comprehensive without a grammar to define how it should be used [67]. Campbell and Musen stated that, in order to provide systematic domain coverage, they would need both a patient-description vocabulary and rules for manipulation of the vocabulary [68]. Rector et al. add an additional requirement: not only is there a grammar for manipulation, but there is concept-specific information about "what is sensible to say" that further limits how concepts can be arranged [43]. Such limitations are needed in order for the vocabulary to support operations such as predictive data entry, natural language processing, and aggregation of patient records; Rector (and others in the Galen Project) simply request that such information be included as part of the vocabulary, in the form of constraints and sanctions [69].

If drawing the line between concept and context can become difficult [41], drawing the line between the vocabulary and the application becomes even more so. After all, the ultimate context for controlled medical vocabulary concepts is some external form such as a patient record. Coping with such contexts may be easier if such contexts are modeled in the vocabulary [70]. A schematic of how such contexts fit together is shown in Fig. 2. The figure differentiates between levels of concept interaction: what's needed to define the concepts, what's desired to show expressivity of the vocabulary, and how such expressiveness is channeled for recording purposes (e.g., in a patient record).

Of course, patient records vary a great deal from institution to institution

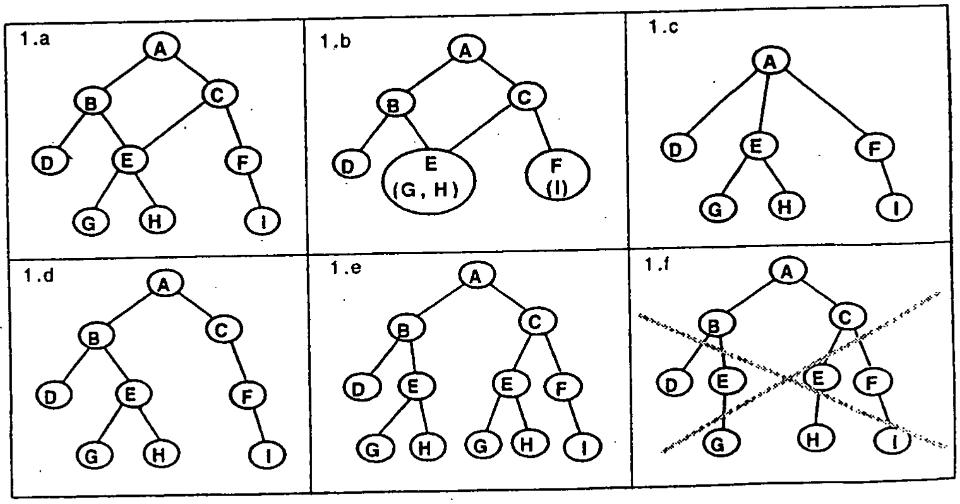


Fig. 1 Multiple views of a polyhierarchy. a) Internal arrangements of nine concepts in a polyhierarchy, where E has two parents; b) Hierarchy has been collapsed so that specific concepts serve as synonyms of their more general parents; c) Intermediate levels in the hierarchy have been hidden; d) Conversion to a strict hierarchy; e) Strict hierarchy with multiple contexts for term E; f) Multiple contexts for E are shown, but are inconsistent (different children).

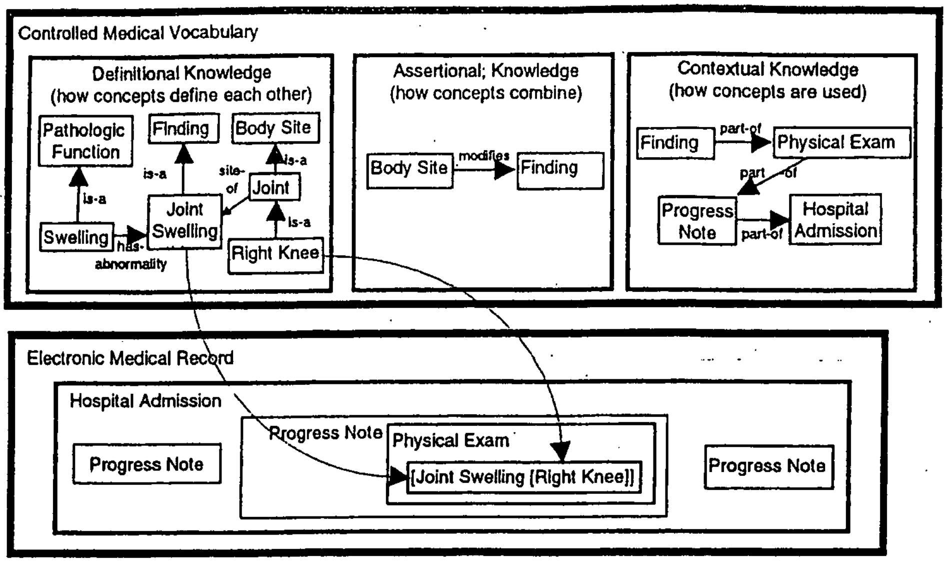


Fig. 2 Definitional, assertional, and contextual information in the vocabulary showing how concepts can be combined and where they will appear in a clinical record.

and, if we have difficulty standardizing on a vocabulary, what hope is there for standardizing on a record structure? One possible solution is to view the recording of patient information from an "event" standpoint, where each event is constitutes some action, including the recording of data, occurring during an episode of care which, in turn occurs as part of a patient encounter [71, 72]. These add more levels to the organization of concepts in contexts, but can be easily modeled in the vocabulary, as in Fig. 2.

2.11 Evolve Gracefully

It is an inescapable fact that controlled vocabularies need to change with time. Even if there were a perfect vocabulary that "got it right the first time", the vocabulary would have to change with the evolution of medical knowledge. All too often, however, vocabularies change in ways that are for the convenience of the creators but wreak havoc with the users [32]. For example, if the name of a concept is changed in such a way as to alter its meaning, what happens to the ability to aggregate patient data that are coded before and after the change? An important desideratum is that those charged with maintaining the vocabulary must accommodate graceful evolution of their content and structure. This can be accomplished through clear, detailed descriptions of what changes occur and why [73], so that good reasons for change (such as simple addition, refinement, precoordination, disambiguation, obsolescence, discovered redundancy, and minor name changes) can be understood and bad reasons (such as redundancy, major name changes, code reuse, and chang codes) can be avoided.[74]

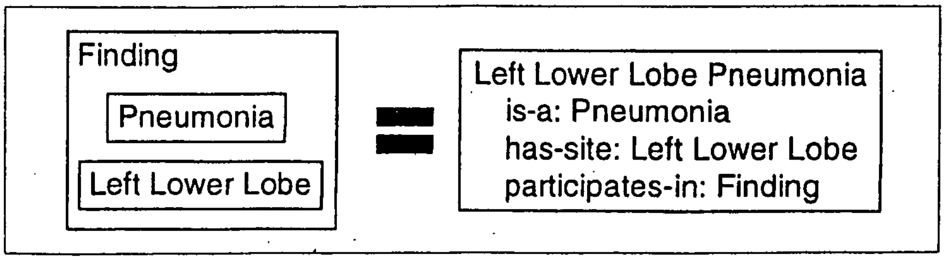


Fig. 3 Interchangability of redundant data representations. The structure on the left depicts the post coordination of a disease concept (Pneumonia) and a body location (Left Lower Lobe) to create a finding in an electronic medical record. The structure on the right shows a precoordinated term for the same finding (Left Lower Lobe Pneumonia). Because this latter term includes formal, structured definitional information (depicted by the is-a, has-site, and participates-in attributes), it is possible to recognize, in an automated way, that data coded in these two different ways are equivalent.

2.12 Recognize Redundancy

In controlled vocabulary parlanc redundancy is the condition in whi the same information can be state in two different ways. Synonymy is type of redundancy which is desirable: helps people recognize the terms th associate with a particular concept an since the synonyms map to the sam concept (by definition), then the codir of the information is not redundan On the other hand, the ability to cod information in multiple ways is genera ly to be avoided. However, such redur dancy may be inevitable in a goo expressive vocabulary.

Consider an application in which th user records a coded problem list. Fo any given concept the user might wis to record, there is always the possibilit that the user desires a more specifi form than is available in the vocabular A good application will allow the use to add more detail to the code problem, either through the addition o a coded modifier, through the use o unconstrained text, or perhaps a combination of both. For example, if a patien has a pneumonia in the lower lobe o the left lung, but the vocabulary doe not have such a concept, the user migh select the coded concept "Pneumonia' and add the modifier "Left Lowe Lobe". Suppose that, a year later, th vocabulary adds the concept "Lef Lower Lobe Pneumonia". Now, there are two ways to code the concept – the old and the new. Even if we were to somehow prevent the old method from being used, we still have old data coded that way.

As vocabularies evolve, gracefully or not, they will begin to include this kind of redundancy. Rather than pretend it does not happen, we should embrace the diversity it represents while, at the same time, provide a mechanism by which we can recognize redundancy and perhaps render it transparent. In the example above, if I were to ask for all patients with "Left Lower Lobe Pneumonia", I could retrieve the ones coded with the specific concept and those coded with a combination of concepts. Such recognition is possible if we have paid sufficient attention to two other desiderata: formal definitions and context representation. If we know, from context representation, that the disease concept "Pneumonia" can appear in a medical record together with an anatomical concept, such as "Left Lower Lobe", and the definition of "Left Lower Lobe Pneumonia" includes named relationships to the concepts "Pneumonia" and "Left Lower Lobe" ("is a" and "site of" relationships, respectively), sufficient information exists to allow us to determine that the representation of the new concept in the vocabulary is equivalent to the collection of concepts appearing in the patient database (see Fig. 3).

3. Discussion

The intense focus previously directed at such issues as medical knowledge representation and patient care data models is now being redirected to the issue of developing and maintaining shareable, multipurpose, high-quality vocabularies. Shareability of vocabulary has become important as system builders realize they must rely on vocabulary builders to help them meet the needs of representing large sets of clinical terms. The multipurpose nature of vocabularies refers to their ability to be used to record data for one purpose (such as direct patient care) and then be used for reasoning about the data (such as automated decision support), usually through a variety of views or abstractions of the specific codes used in data capture. Even the ability to support browsing of vocabularies remains problematic [75]. "High quality" has been difficult to define, but generally means that the vocabulary approaches completeness, is well organized, and has terms whose meanings are clear. The above list of desiderata for shareable, multipurpose controlled vocabularies reflect one person's view of the necessary priorities; however, they are based on personal experience with attempts to adopt vocabularies [76-79] and gleaned from the reported experiences of others. The solutions necessary to meet the above list of desiderata vary from technical to political, from simple adoption to basic shifts in philosophy, and from those currently in use to areas ripe for research.

Developers of controlled vocabularies are recognizing that their products are in demand for multiple purposes and, as such, they must address a variety of needs that go beyond those included for the vocabulary's original purpose [80]. The simple solution of "add more terms until they're happy" is not satisfying vocabulary users; they want content, but they want more. They want information about the terms, so they know what they mean and how to use them. They also want this information to supplement the knowledge they create for their own purposes. These purposes are as diverse as natural language processing, predictive data entry, automated decision support, indexing, clinical research, and even the maintenance of vocabularies themselves.

Simple, technical solutions are at hand for some characteristics, and are already being adopted. For example, using nonsemantic concept identifiers and allowing polyhierarchies are straightforward. The systematic solution for some others, such as multiple granularities and multiple consistent views will require more thought, but generally should be tractable. Allowing graceful evolution and recognized redundancy are still areas for research, with some promising findings. For example, systematic approaches for vocabulary updates are being discussed to support evolution [73], while conceptual graphs provide a mechanism for transforming between different synonymous (i.e., redundant) arrangements of associated concepts [54].

Some of the desiderata will require fundamental philosophical shifts. For example, decisions to have a truly concept-oriented vocabulary and avoid the dreaded "NEC" terms are simple ones, but can not be taken lightly. Some of these decisions, such as formal definitions and representing context, will also require significant development effort to make them a reality. Several developers describe commitment to these goals, and one group has actually provided formal, computer-manipulable definitions of their concepts [81, 82]. But the amount of work remains formidable. Finding ways to share the burden of vocabulary design and construction will be challenging [83], but some proaches seem promising [59]. Fin ways to coordinate content deve ment and maintenance among mul groups will require sophisticated ap aches [60]. Despite their perce infancy [84], the currently avail standards should be the starting p for new efforts [85].

Predictions may not be difficul make, given the current direction which standards development is ceeding. It is likely that vocabula will become concept-oriented, u nonsemantic identifiers and contair semantic information in the form semantic network, including mult hierarchies. Development of a stand notation for the semantic informa may take some time, but the concep graph seems to be a popular candid Maintenance of vocabularies will ev tually settle down into some form wh is convenient for users and conc permanence will become the no Still unclear is whether the seman definitional information provided developers will be minimal, comple or somewhere in between. Some of other desiderata, such as context rep sentation, multiple consistent vie and recognition of redundancy probably be late in coming. Howe the knowledge and structure provic with the vocabulary will at least fac tate development of implementatic specific solutions which have not he tofore been possible.

4. Conclusions

This list of desiderata is not intend to be complete; rather, it is a partial l which can serve to initiate discussi about additional characteristics need to make controlled vocabularies sha ble and reusable. The reader should n infer that vocabulary developers are n addressing these issues. In fact, the same developers were the sources f many of the ideas listed here. As result, vocabularies are undergoi their next molt. Current trends seem indicate that this one will be a tr metamorphosis, as lists change to mul ple hierarchies, informal descripti information changes to formal defi tional and assertional information, ar attention is given not just to the expansion of content, but to structural and representational issues.

Acknowledgments

The ideas in this paper have been based on the medical informatics literature and synthesized from years of work and conversations with many researchers. In particular, the author's collaborators on the UMLS project, the Canon group, and the InterMed Collaboratory can be credited with helping to precipitate these thoughts. In addition, since this paper has been prepared as an accompaniment to a presentation at the IMIA Working Group 6 conference, it is hoped that additional desiderata will be put forth at the meeting which will find their way into the final version of this manuscript. This work has been supported in part by the National Library of Medicine.

REFERENCES

1. Wong ET, Pryor TA, Huff SM, Haug PJ, Warner HR. Interfacing a stand-alone diagnostik expert system with a hospital information system. Comput Biomed Res 1994; 27: 116-29.
2. Masys DR. Of codes and keywords: standards for biomedical nomenclature. Acad Med 1990; 65: 627-9.
3. Cimino JJ. Coding systems in health care. In van Bemmel JH, McCray AT, eds.: Yearbook of Medical Informatics, International Medical Informatics Association, Rotterdam, 1995: 71-85. Reprinted in Methods of Information in Medicine 1996; 35 (4/5): 273-84.
4. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The Content Coverage of Clinical Classifications. JAMIA 1996; 3: 224-33.
5. Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes; completeness, taxonomy, mapping, definitions and clarity. Journal of the American Medical Informatics Association 1997; 4: 238-51.
6. Conference Summary Report: Moving Toward International Standards in Primary Care Informatics: Clinical Vocabulary. New Orleans, November, 1995. United States Department of Health and Human Service 1996.
7. Anderson J. The computer: medical vocabulary and information. British Medical Bulletin 1968; 24 (3): 194-8.
8. Bates JAV. Preparation of clinical data for computers. British Medical Bulletin 1968; 24 (3): 199-205.
9. Howell RW, Loy RM. Disease coding by computer: the "fruit machine" method. British Journal of Preventive and Social Medicine 1968; 22: 178-81.
10. Cimino JJ. Controlled medical vocabulary construction: methods from the Canon Group. JAMIA. 1994; 1: 296-7.
11. Ozbolt JG, Fruchtnicht JN, Hayden JR. Toward data standards for nursing information. JAMIA. 1994; 1: 175-85.
12. McCloskey J, Bulechek G. Letter: Toward data standards for nursing information. JAMIA 1994; 1: 469-71.
13. Ozbolt JG, Fruchtnicht JN, Hayden JR. Letter: Toward data standards for nursing information. JAMIA 1994; 1: 471-2.
14. Humphreys BL. Comment: Toward data standards for nursing information. JAMIA 1994; 1: 472-4.
15. Goossen WTF, Epping PJMM, Abraham IL. Classification systems in nursing: formalizing nursing knowledge and implications for nursing information systems. Meth Inform Med. 1996; 35: 59-71.
16. Gabrieli ER. Computerizing text from office records. MD Comput. 1987; 4: 44-9, 56.
17. Côté RA, Robboy S. Progress in medical information management – Systematized nomenclature of medicine (SNOMED). JAMA 1980; 243: 756-62.
18. Evans DA, Rothwell DJ, Monarch IA, Lefferts RG, Côté RA. Towards representations for medical concepts. Med Decis Making 1991; 11 (suppl): S102-S108.
19. Musen MA, Weickert KE, Miller ET, Campbell KE, Fagan LM. Development of a controlled medical terminology: knowledge acquisition and knowledge representation. Meth Inform Med 1995; 34: 85-95.
20. Moehr JR, Kluge EH, Patel VL. Advanced patient information systems and medical concept representation. In Greenes RA, Peterson HE, Protti DJ, eds. Proceedings of the Eight World Congress on Medical Informatics (MEDINFO '95), Vancouver, British Columbia, Canada. Healthcare Computing & Communications (Canada); 1995: 95-9.
21. Evans DA. Pragmatically-structured, lexical-semantic knowledge bases for a Unified Medical Language System. In Greenes RA, ed. Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care (SCAMC). Washington, DC; New York: IEEE Computer Society Press 1988: 169-73.
22. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Meth Inform Med 1993; 32: 281-91.
23. Volot F, Zweigenbaum P, Bachimont B, Ben Said M, Bouaud J, Fieschi M, Boisvieux JF. Structuration and acquisition of medical knowledge using UMLS in the conceptual graph formalism. In Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care (SCAMC). Washington, DC, 1993; McGraw-Hill, New York; 1994: 710-14.
24. Evans DA, Cimino JJ, Hersh WR, Huff SM, Bell DS. Toward a medical-concept representation language. JAMIA 1994; 1: 207-17.
25. Rassinoux AM, Miller RA, Baud RH, Scherrer JR. Modeling principles for QMR medical findings. In Cimino JJ, ed. Proceedings of the AMIA Annual Fall Symposium (Formerly SCAMC). Washington, DC, Philadelphia: Hanley and Belfus 1996: 264-98.
26. Henkind SJ, Benis AM, Teichholz LE. Quantification as a means to increase the utility of nomenclature-classification systems. In Salmon R, Blum B, Jürgensen, eds. Proceedings of the Seventh World Congress on Medical Informatics (MED-INFO 86). North-Holland (Amsterdam); 198 858-61.
27. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to th maintenance of a large controlled medica terminology. JAMIA 1994; 1: 35-50.
28. Blois MS. The effect of hierarchy on th encoding of meaning. In Orthner HF, ed Proceedings of the Tenth Annual Symposium on Computer Applications in Medica Care (SCAMC). Washington, DC; IEEE Computer Society Press, New York; 1986:73
29. Moorman PW, van Ginneken AM, van der Lei J, van Bemmel JH. A model for structured data entry based on explicit descriptiona knowledge. Meth Inform Med 1994; 33 454-63.
30. van Ginneken AM, van der Lei J, Moorman PW. Towards unambiguous representation of patient data. In Frisse ME, ed. Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care (SCAMC). Baltimore, MD, 1992; McGraw-Hill, New York; 1993: 69-73.
31. Cimino JJ, Hripcsak G, Johnson SB, Clayton PD. Designing an introspective, controlled medical vocabulary. In Kingsland LC, ed. Proceedings of the Thirteenth Annual Symposium on Computer Applications in Medical Care (SCAMC). New York, Washington, DC: IEEE Computer Society Press 1989: 513-8.
32. Cimino JJ. Formal descriptions and adaptive mechanisms for changes in controlled medical vocabularies. Meth Inform Med. 1996; 35: 202-10.
33. Forman BH, Cimino JJ, Johnson SB, Sengupta S, Sideli R, Clayton P. Applying a controlled medical terminology to a distributed production clinical information system. In Gardner RM, ed. Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care (SCAMC). Philadelphia, New Orleans, LA: Hanley and Belfus 1995: 421-5.
34. Bernauer J, Franz M, Schoop M, Schoop D, Pretschner DP. The compositional approach for representing medical concept systems. In Greenes RA, Peterson HE, Protti DJ, eds. Proceedings of the Eight World Congress on Medical Informatics (MEDINFO '95), Vancouver, British Columbia, Canada. Healthcare Computing & Communications (Canada) 1995: 70-4.
35. Dunham GS, Henson DE, Pacak MG. Three solutions to problems of categorized medical nomenclatures. Meth Inform Med 1984; 23: 87-95.
36. Baud R, Lovis C, Alpay L, Rassinoux AM, Nowlan A, Rector A. Modeling for natural language understanding. In Safran C, ed. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care (SCAMC). New York, Washington, DC: 1993; New York: McGraw-Hill 1994: 289-93.
37. Zweigenbaum P, Bachimont B, Bouaud J, Charlet J, Boisvieux JF. Issues in the structuring and acquisition of an ontology for medical language understanding. Meth Inform Med 1995; 34: 15-24.
38. Masarie Jr. FE, Miller RA, Bouhaddou O, Giuse NB, Warner, HR. An interlingua for

electronic interchange of medical information: using frames to map between clinical vocabularies. Computers in Biomedical Research 1991; 24 (4): 379-400.

39. Gabrieli ER. Interface problems between medicine and computers. In Cohen GS, ed. Proceedings of the Eighth Annual Symposium on Computer Applications in Medical Care (SCAMC). New York, Washington, DC: IEEE Computer Society Press 1984: 93-5.

40. Barr CE, Komorowski HJ, Pattison-Gordon E, Greenes RA. Conceptual modeling for the Unified Medical Language System. In Greenes RA, ed. Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care (SCAMC). New York, Washington, DC: IEEE Computer Society Press 1988: 148-51.

41. Haimowitz IJ, Patil RS, Szolovitz P. Representing medical knowledge in a terminological language is difficult. In Greenes RA, ed. Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care (SCAMC). New York, Washington, DC: IEEE Computer Society Press 1988: 101-5.

42. Rothwell DJ, Côté RA. Optimizing the structure of a standardized vocabulary – the SNOMED model. In Miller RA, ed. Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care (SCAMC). New York, Washington, DC: IEEE Computer Society Press 1990: 181-4.

43. Rector AL, Nowlan WA, Kay S. Conceptual knowledge: the core of medical information systems. In Lun KC, Deguolet P, Piemme TE, Rienhoff O, eds. Proceedings of the Seventh World Congress on Medical Informatics (MEDINFO '92), Geneva. North-Holland (Amsterdam) 1992: 1420-6.

44. Campbell KE, Das AK, Musen MA. A logical foundation for representation of clinical data. JAMIA 1994; 1: 218-32.

45. Bernauer J. Subsumption principles underlying medical concept systems and their formal reconstruction. In Ozbolt JG, ed. Proceedings of the Eighteenth Annual Symposium on Computer Applications in Medical Care (SCAMC). Washington, DC; Hanley and Belfus, Philadelphia; 1994: 140-4.

46. Smart JF, Roux M. A model for medical knowledge representation application to the analysis of descriptive pathology reports. Meth Inform Med 1995; 34: 352-60.

47. Kiuchi T, Ohashi Y, Sato H, Kaihara S. Methodology for the construction of a disease nomenclature and classification system for clinical use. Meth Inform Med 1995; 34: 511-7.

48. Rector AL. Thesauri and formal classifications: terminologies for people and machines. In Chute CG, ed. IMIA Working Group 6 Conference on Natural Language and Medical Concept Representation, Jacksonville, Florida, January 1997: 183-95.

49. Brown PJB, O'Neil M, Price C. Semantic representation of disorders in Version 3 of the Read Codes. In Chute CG, ed. IMIA Working Group 6 Conference on Natural Language and Medical Concept Representation, Jacksonville, Florida: January 1997: 209-14.

50. Price C, O'Neil M, Bentley TE, Brown PJB. Exploring the ontology of surgical procedures in the Read Thesaurus. In Chute CG, ed. IMIA Working Group 6 Conference on Natural Language and Medical Concept Representation, Jacksonville, Florida, January 1997: 215-21.

51. Bernauer J. Formal classification of medical concept descriptions: graph oriented operators. In Chute CG, ed. IMIA Working Group 6 Conference on Natural Language and Medical Concept Representation, Jacksonville, Florida, January 1997: 109-15.

52. Rossi Mori A, Consorti F, Galeazzi E. Standards to support development of terminological systems for healthcare telematics. In Chute CG, ed. IMIA Working Group 6 Conference on Natural Language and Medical Concept Representation, Jacksonville, Florida, January 1997: 131-45.

53. Bernauer J. Conceptual graphs as a operational model for descriptive findings. In Clayton PD, ed. Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care (SCAMC). Washington, DC, 1991; McGraw-Hill, New York; 1992: 214-8.

54. Campbell KE, Musen MA. Representation of clinical data using SNOMED III and conceptual graphs. In Frisse ME, ed. Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care (SCAMC). Baltimore, MD, 1992; McGraw-Hill, New York; 1993: 354-8.

55. Baud RH, Rassinoux AM, Wagner JC, Lovis C, Juge C, Alpay LL, Michel PA, Degoulet P, Scherrer JR. Representing clinical narratives using conceptual graphs. Meth Inform Med 1995; 34: 176-86.

56. Robé PF de V, Flier FJ, Zanstra PE. Health classifications and terminological modeling. In Chute CG, ed. IMIA Working Group 6 Conference on Natural Language and Medical Concept Representation, Jacksonville, Florida, January 1997: 223-4.

57. Cimino JJ, Barnett GO. Automatic knowledge acquisition from MEDLINE. Meth Inform Med 1993; 32: 120-30.

58. Cimino JJ, Johnson SB, Hripcsak G, Hill CL, Clayton PD. Managing Vocabulary for a Centralized Clinical System. In Greenes RA, Peterson HE, Protti DJ, eds. Proceedings of the Eight World Congress on Medical Informatics (MEDINFO '95), Vancouver, British Columbia, Canada. Healthcare Computing & Communications (Canada); 1995: 117-20.

59. Friedman C, Huff SM, Hersh WR, Pattison-Gordon E, Cimino JJ. The Canon group's effort: working toward a merged model. JAMIA 1995; 2: 4-18.

60. Campbell KE, Cohn SP, Chute CG, Rennels G, Shortliffe EH. Gálapagos: computer-based support for evolution of a convergent medical terminology. In Cimino JJ, ed. Proceedings of the AMIA Annual Fall Symposium (Formerly SCAMC). Washington, DC; Hanley and Belfus, Philadelphia; 1996: 269-73.

61. Campbell KE, Cohn SP, Chute CG, Shortliffe EH, Rennels G. Scaleable methodologies for distributed development of logic-based convergent medical terminology. In Chute CG, ed. IMIA Working Group 6 Conference
on Natural Language and Medical Conc
Representation, Jacksonville, Florida, J
uary 1997: 243-56.

62. Musen MA. Knowledge sharing and reι
Comput Biomed Res 1992; 25: 435-67.

63. Blois MS. Medicine and the nature of ve
cal reasoning. New Engl J Med 1988; 3
847-51.

64. Kong A, Barnett GO, Mosteller F, Youtz
How medical professionals evaluate expr
sions of probability. New Engl J Med 19
315: 740-4.

65. Henry SB, Mead CN. Standardized nursi
classification systems; necessary, but not s
ficient, for representing what nurses do.
Cimino JJ, ed. Proceedings of the AM
Annual Fall Symposium (Forme
SCAMC). Philadelphia, Washington, D
Hanley and Belfus 1996: 145-9.

66. Huff SM, Rocha RA, Solbrig HR, Barr
MW, Schank SP, Smith M. Linking a medic
vocabulary to a clinical data model usi
Abstract Syntax Notation 1. In Chute C
ed. IMIA Working Group 6 Conference
Natural Language and Medical Conce
Representation, Jacksonville, Florida, Jan
ary 1997: 225-41.

67. Huff SM, Craig RB, Gould BL, Castag
DL, Smilan RE. Medical data dictionary f
decision support applications. In Stead W
ed. Proceedings of the Eleventh Annu
Symposium on Computer Applications
Medical Care (SCAMC). New York, Wa
hington, DC: IEEE Computer Society Pre
1987: 310-7.

68. Campbell KE, Musen MA. Creation of
systematic domain for medical care: th
need for a comprehensive patient-descri
tion vocabulary. In Lun KC, Deguolet
Piemme TE, Rienhoff O, eds. Proceedings
the Seventh World Congress on Medic
Informatics (MEDINFO '92), Genev
North-Holland (Amsterdam); 1992: 1437-4

69. Rector AL, Glowinski AJ, Nowlan WA
Rossi-Mori A. Medical-concept models an
medical records: an approach based o
GALEN and PEN&PAD. JAMIA 1995;
19-35.

70. Prokosh HU, Amiri F, Krause D, Neek C
Dudeck J. A semantic network model fo
the medical record rheumatology clinic. I
Greenes RA, Peterson HE, Protti DJ, eds
Proceedings of the Eight World Congres
on Medical Informatics (MEDINFO '95
Vancouver, British Columbia, Canada
Healthcare Computing & Communication
(Canada); 1995: 240-4.

71. Huff SM, Rocha RA, Bray BE, Warner HR
Haug PJ. An event model of medical infor
mation representation. JAMIA 1995;
116-34.

72. Johnson SB, Friedman C, Cimino JJ, Clar
T, Hripcsak G, Clayton PD. Conceptual dat
model for a central patient database. I
Clayton PD, ed. Proceedings of the Fifteent
Annual Symposium on Computer Applica
tions in Medical Care (SCAMC). Washing
ton, DC 1991; McGraw-Hill, New Yor
1992: 381-5.

73. Tuttle MS, Nelson SJ. A poor precedent
Meth Inform Med. 1996; 35: 211-7.

74. Cimino JJ, Clayton PD. Coping with chang-
ing controlled vocabularies. In Ozbolt JG

ed.: Proceedings of the Eighteenth Annual Symposium on Computer Applications in Medical Care; New York, Washington, DC: November, McGraw-Hill, 1994: 135-9.

75. Hripcsak G, Allen B, Cimino JJ, Lee R. Access to data: comparing Access Med with Query by Review. JAMIA. 1996; 3: 288-99.

76. Cimino JJ. Representation of clinical laboratory terminology in the Unified Medical Language System. In Clayton PD, ed. Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care (SCAMC). Washington, DC, 1991; McGraw-Hill, New York; 1992: 199-203.

77. Cimino JJ, Sideli RV. Using the UMLS to bring the library to the bedside. Medical Decision Making. 1991; 11 (Suppl): S116-S120.

78. Cimino JJ, Barrows RC, Allen B. Adapting ICD9-CM for clinical decision support (abstract). In Musen MA, ed. Proceedings of the 1992 Spring Congress of the American Medical Informatics Association; Portland, Oregon; May, 1992. The Association, Bethesda, MD; 1992: 34.

79. Cimino JJ. Use of the Unified Medical Language System in patient care at the Columbia-Presbyterian Medical Center. Meth Inform Med. 1995; 34: 158-64.

80. Schulz EB, Price CS, Brown PJB. Symbolic anatomic knowledge representation in the Read Codes Version 3: structure and application. Journal of the American Medical Association. 1997; 4: 38-48.

81. Forrey AW, McDonald CJ, DeMoor G, Huff SM, Leaville D, Leland D, Fiers T, Charles L, Griffin B, Stalling F, Tullis A, Hutchins K, Baenziger J. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. Clinical Chemistry. 1996: 42: 81-90.

82. Rocha RA, Huff SM. Coupling vocabularies and data structures: lessons from LOINC. In Cimino JJ, ed. Proceedings of the AMIA Annual Fall Symposium (Formerly SCAMC). Washington, DC; Hanley & Belfus, Philadelphia; 1996: 90-4.

83. Tuttle MS. The position of the Canon gro a reality check. JAMIA 1994; 1: 298-9.

84. United States General Accounting Offi Automated Medical Records: Leaders. Needed to Expedite Standards Devel ment: report to the Chairman/Committee Governmental Affairs, U.S. Senate. Washi ton, DC: USGAO/IMTEC-93-17; April 19

85. Board of Directors of the American Medi Informatics Association. Standards medical identifiers, codes, and messag needed to create an efficient compute stored medical record. JAMIA 1994; 1: 1-

Address of the author:
James J. Cimino, M.D.,
Department of Medical Informatics,
Columbia University College of Physicians
and Surgeons,
161 Fort Washington Avenue,
New York, New York 10032, USA
E-mail: James.Cimino@columbia.edu

Exhibit G

MEDINFO 98
B. Cesnik et al. (Eds)
Amsterdam: IOS Press

The Role of Compositionality in Standardized Problem List Generation

P.L. Elkin[a], M. Tuttle[b], K. Keck[b], K. Campbell[c], G. Atkin[a], C.G. Chute[a].

[a]*Mayo Foundation*
[b]*Lexical Technology, Inc.*
[c]*Stanford University*

Abstract

Compositionality is the ability of a Vocabulary System to record non-atomic strings. In this manuscript we define the types of composition, which can occur. We will then propose methods for both server based and client-based composition. We will differentiate the terms Pre-Coordination, Post-Coordination, and Ûser-Directed Coordination. A simple grammar for the recording of terms with concept level identification will be presented, with examples from the Unified Medical Language System's (UMLS) Metathesaurus. We present an implementation of a Window's NT^tm based client application and a remote Internet Based Vocabulary Server, which makes use of this method of compositionality. Finally we will suggest a research agenda which we believe is necessary to move forward toward a more complete understanding of compositionality. This work has the promise of paving the way toward a robust and complete Problem List Entry Tool.

Keywords

Structured Entry; Controlled Medical Vocabularies; Documentation; Compositionality

Introduction

Vocabulary construction and organization is seen as an essential part of a functional Electronic Health Record[1]. Concept level understanding of our day to day clinical practice will enable more accurate and more available outcomes research, evidence based medicine, and effective cost management of medicine without a decline in service. This promise is hampered by the lack of a robust clinically relevant large-scale vocabulary, with a structure which supports synonymy, multiple ontologies, semantic relationships, and compositionality. As we move toward a greater understanding of the relationships between terms, workers are striving to determine the optimal level of granularity for the terms in these Vocabularies. One solution would be to separate the truly atomic terms and their ontology from the compositions and their relationships. This multiaxial schema for Vocabulary design is clearly controversial. An example of this type of construction would be "Coronary Artery Disease (CAD) Status Post CABG", here we have multiple atomic concepts. On first cut, the Coronary Artery Disease can be separated from the s/p CABG. This is only possible, if there exists a mechanism for reconstruction. This is clinically very important because the patient with CAD s/p CABG is clearly a different presentation than a patient with CAD without a history of prior cardiac surgery. More controversial is the corollary, that the construction of Coronary Artery and Atherosclerotic Vascular Disease, should be an equivalent concept to CAD.

Although we may wish to say many things about CAD as a unit, there are still more granular ways to represent the same concepts. This similarity can be seen in many other constructions, for example the combination of "Large Bowel" and "Neoplasm, malignant" is equivalent to "Colon Cancer". This is particularly important for billing systems where the code for "colon cancer" might have a different ICD9-CM code than the two terms "large bowel" and "neoplasm, malignant". One challenge in the development of a canonical vocabulary is to eliminate redundancy. Composition, while powerful, is also a source of considerable redundancy.

If composition causes such angst, why do it? Why not ignore this functionality? The answer became clear to our group during a recent Usability Trial conducted at the Mayo Clinic[2]. Users demand the ability to form problem statements that represent the concepts of their practice. We do not and can not anticipate everything a clinician might wish to say about a patient. Thus without fully functional Natural Language Processing, we can not represent clinical medicine completely within a well-formed controlled vocabulary. One solution is compositionality. All of these complex and varied statements that clinicians make regarding their patients are derived from a manageable number of atomic concepts (estimated to fall somewhere between (20,000 and 1,000,000) [3,4,5].

Glossary

Atomic Concept: A notion represented by language, which identifies one idea. Such an entity can not be broken into parts without the loss of meaning.

Example: In the UMLS Metathesaurus, Colon is a synonym for Large Bowel and Cancer is a synonym for Neoplasm, Malignant. Whereas Colon Cancer is non-atomic as it can be broken down into "Large Bowel" and "Neoplasm, malignant". Each of these two more atomic terms has a separate and unique Concept Unique Identifier (CUI).

Pre-coordinated Concept: A notion represented by language, which identifies one idea. Such an entity can be broken into parts without loss of meaning when the atomic concepts are examined in combination. These are terms, which are considered single concepts within the host vocabulary.

Example: Colon Cancer is non-atomic, however it has a single CUI, which means to the Metathesaurus that it represents a "single" concept. It has the same status in the vocabulary as the site "Large Bowel" and the diagnosis "Neoplasm, malignant".

Post-coordinated Concepts: A notion represented by language and a set of codes (concept level identifiers), which identifies one idea. This is the attempt of a system to construct a set of concepts from within a controlled vocabulary to more completely represent a user's query.

Example: The concept "Status-Post CABG" is not a unique term within the UMLS Metathesaurus. It represents a clinical concept that some patient has already had heart surgery. As it can not be represented by a single CUI, to fully capture the intended meaning a system would need to build a representation from multiple CUIs or lose information to free text.

User-Directed Coordination of Concepts: A notion represented by language and a set of codes (concept level identifiers), which identifies one idea. The User chooses this set of concepts, usually via a Graphical User Interface, and usually we envision that this would occur at the point-of-care. This is the attempt of a User to represent a clinical concept using a set of concepts, whether they are atomic, pre-coordinated, or post-coordinated concepts. The clinician's focus is to most fully capture the meaning that they wish to record regarding their patients.

Example: A GUI, which enables users to combine concepts in a meaningful way. This in our view implies a robust representational schema. Such a schema would facilitate an understanding of these compound structures and their relative locality within the canonical vocabulary. These structures should be non-redundant and should facilitate vocabulary maintenance.

Methods of Composition

Vocabulary Based Strategies

Natural Language Processing is a complex computational task. Systems capable of understanding free speech are not presently available, however many useful and reliable tools have been developed.[3,6,7,8,9] Harnessing the information inherent in the input string is essential to providing the sort of useful service, which the busy clinician demands. We advocate parsing the input strings into Main Concepts, Qualifiers and Modifiers and knowing the types of relationships that classes of Qualifiers can have with main concepts, and each other, we can provide better post-coordination of matched compound concepts (multiple Concept Unique Identifiers). Qualifiers are terms, which change the meaning of a term in a temporal or administrative sense, as opposed to a clinical sense (i.e. "History of", "Status/Post", "Recurrent", "Rule-Out", etc.).[10] These compound concepts need to be linked / built-up in a meaningful and useful manner. Utilizing as much as possible the clues which we are given from the input string is an important mechanism for accomplishing this task.

An example of would be the input statement:

History of Benign Prostatic Hypertrophy (BPH)
11111111 222222222222222222222222

Status/Post Transurethral Resection of the Prostate (TURP)

333333333 444444444444444444444444444444

Here there are represented four unique concepts. We know that "History of" and "Status/Post" are both qualifiers, and that BPH and TURP are both undifferentiated problems. The term "History of" can relate to just BPH or to both BPH and TURP. The term "Status/Post" always acts on the next concept or set of concepts, and therefore must relate to TURP (S/P TURP). Hence this expression could be interpreted as either (represented in ASN.1):

1. Concept {{Qualifier "Concept 1", Base-Concept {name "Concept 2}}, {Qualifier "Concept 3", Base-Concept {name "Concept 4"}}} **or**
2. Concept {{Qualifier "Concept 1", Base-Concept {Concept {Base-Concept {name "Concept 2"}}, {qualifier "Concept 3", Base-Concept {name "Concept 4"}}}}}}

In the first example concept one qualifies just concept two, and in the second it qualifies concepts two and four, whereas concept three always qualifies only concept 4.

Server Based Strategies

Distributed computing theory recognizes that Servers are efficient at handling large amounts of information. They however are not good at handling process intensive tasks, by virtue of the fact that many users will in all likelihood be using the server simultaneously. Therefore we recommend pushing process intense tasks to the client when feasible, given the availability of relatively cheap cycle time.

1. Vocabulary Storage and Retrieval

The capability of massive data storage and retrieval with buffering of indices, which can be accessed by multiple simultaneous processes, makes server side retrievals fast and efficient. Maintenance and updating of the vocabulary need be done in only one place for all users to benefit. Better version consistency can be maintained.

2. Universal Unique Identifiers for Compound Concept Unique Identifiers

We will never want to maintain a concept in the base vocabulary for every compound concept that a user may want to express. For example "History of BPH s/p TURP" does not make most workers lists of atomic concepts. On the other hand a clinician may very well wish to make this statement regarding one of their patients. Each and every time such a reference is used, we would want to capture its meaning and if another clinician wrote "Hx of BPH two years after a TURP" it would be nice if a system could recognize these as being related to the same set of concepts. This requires that the server serve up the same identifier not only for unique concepts but also for unique compound concepts.

3. Making the most out of your retrieval list

We make use of simple rules of composition, which uses an ontology of qualifiers which can be combined in selected ways with other concepts selected by UMLS semantic type (i.e. Problem, Disorder, etc.). Multiple qualifiers can be combined with multiple other concepts to provide a short list of retrievals, which a clinician might choose as their problem list. entry. These post-coordinated terms are presented at the top of the retrieval list, but with no indication 'that they differ from any other term presented on the retrieval list (e.g. Unique atomic concepts, Unique pre-coordinated concepts).

Client Based Strategies:

Client side applications have two powerful and distinct advantages over the server. First, the client is blessed with excess cycle time, and second it can interact with the user tapping the tremendous discriminating power of the human mind. The client side embodiment of compositionality makes use of a Graphical User Interface to provide rapid access to relevant information, by the use of relatively simple heuristics. First, we believe that most commonly relevant terms will be found to be matches or synonyms of the user's input. Second, we believe that related terms can stimulate useful thought on the part of the user. Third, we assert that definitions for terms will yield enhanced confidence in clinicians choices of suggested terminology. Fourth, we believe that the ability to mix and match combinations of concepts from either the suggested match list or the related terms will be valuable. Finally we assert that this will always be an incomplete solution. Since the user's thought process is an ever-evolving set of concepts and relations, the user may easily be prompted by items on the return list to enter even more specific ideas. Therefore we allow the user the ability to perform sub-searches within their original retrieval, to help to decrease the cognitive overhead and disorientation which might accompany having to re-think their query.

1. Suggested matches

This is the primary retrieval list, it is the server's best guesses as to concepts and compound concepts that the system recognizes which may approximate the user's query. This list is the closest to the entry field and the cursor is focussed on the best match for the query, such that the user need only hit enter to accept.

2. Related matches

These are automatically generated from the best match within the retrieval list. If the user single-clicks on any term in the suggested matches' list, new related terms are generated. These terms are narrower-than and other-related terms from the UMLS. We are currently in the middle of an effort to establish the relationships, which exist within the Mayo Clinic's problem list vocabulary. This effort would provide related problems from our vocabulary to add to those already generated within the UMLS.

3. Definitions

Definitions are displayed for any term which is given the user's focus within the suggested terms list or the related terms list. This provides the user with increased confidence when choosing a term. These definitions are derived where available from within the UMLS.

4. Composition

A. User level composition is a powerful tool, which leverages a wealth of knowledge regarding term inter-relationships, which are not available within today's controlled vocabularies. We harness this capability by allowing the users to combine concepts from both the Suggested and Related terms boxes.

B. Users then have the ability to run a sub-set of nested searches which maintain the focus of their original search, but allows them to add qualifiers and modifiers to better define a central set of concepts. For example if you entered the term "Left forearm Cellulitis", the retrieval list would return "Left forearm" and "Cellulitis." One might compose the combined concept "Cellulitis, Left Forearm." While making this assignment the user might decide that this is an unusually severe case. They can then enter a composition search for "Severe" and when found could add it to the compound concept making "Cellulitis, Left Forearm, Severe."

C. User level composition leverages combinations of arbitrarily large numbers of atomic, pre-coordinated and post-coordinated server suggested and related terms.

Discussion

Functionality:

The Unified Medical Language System's Metathesaurus houses many dozen vocabularies. Among the contributing vocabularies are SNOMED, the Read Codes, ICD9-CM, CPT, and added to this montage are the clinically derived terminologies from the Mayo Clinic and the Beth Israel Hospital. This Metathesaurus holds 150,000 unique concepts and over 400,000 unique strings (synonyms). This combination of atomic and pre-coordinated terms forms the basis for our compositional model, which harnesses the user's clinical acumen and focuses a query, in order to rapidly return a short and relevant retrieval list. Our Windows NT^{tm} based Graphical User Interface provides an intuitive mechanism for finding or building the problem which a clinician wishes to record regarding their patient's condition. This is accomplished by three mechanisms first, we provide a tool which rapidly produces a relevant retrieval list for a users entry. Second, we automatically present the user with related terms, which might better represent the concept, which the clinician wishes to record. Third, we provide a mechanism for the user to create a compound concept, which can be composed of atomic, pre-coordinated or post-coordinated terms. This flexibility optimizes the system's performance in the case where the system finds a term in the vocabulary, which matches the users input or a synonym. However our system allows the user who does not

· user's
'ns list.
choos-
· :e from

verages
...onships,
d vocabu-
ie users to
:· Related

·f nested
al search,
to better
·· entered
... would
: compose
.." While
: this is an
·mposition
l it to the
Forearm,

s of arbi-
and post-

ris houses
cabularies
l added to
; from the
thesaurus
e strings
ordinated
which har-
y, in order
Our Win-
n intuitive
ch a clini-
on. This is
e a tool
users entry.
ted terms,
: clinician
the user to
of atomic,
bility opti-
e system
rs input or .
e does not

find one term in the terminology which covers the entire concept that the clinician wishes to represent, the ability to mix and match concepts in order to form compound concepts which more closely represents the entire concept. Lastly we save an ontology of compound concepts. This allows the same unique compound representation to be identified uniquely, each time it is referenced by a clinician. This provides uniformity of meaning beyond the scope of the base vocabulary.

Usability:

This system was tested in our usability laboratory. We at the Mayo Clinic have a Usability Laboratory which contains a sound proof room and a control booth where a user's actions and words can be recorded along with the output from a computer screen. These studies are performed to evaluate software or systems, prior to their deployment within the Clinic. We constructed eleven scenarios where clinicians interpret clinical cases and then state aloud the diagnoses for which they are looking. We then can record what they enter, the systems response and of course what the user ultimately chooses for entry into their patient's record. Participants are asked how they found the system's usability, speed, reliability, and they are also encouraged to comment on portions of the system which they feel need improvement. [12,13]

Compositionality was found to be an essential part of the problem manager's functionality. The greatest utility stemmed from the addition of modifiers and qualifiers to primary terms. The success of the program hinged more tightly to the interface design than with other components of the system, which require considerably less cognitive overhead from our users. Users were shown two different interfaces one which was complicated and powerful (allowed drag and drop functions to multiple targets, etc.) and a simpler more straightforward interface with the same vocabulary options but fewer computing options. The users found the former completely unacceptable and blamed the controlled vocabulary itself, while with the latter they were able to successfully navigate the vocabulary building compositional constructs which satisfied their clinical intent. It was also clear that users expressed no trepidation when selecting compound terms. Users found the rapid access to common qualifiers and modifiers (e.g. Acute, Left, Severe) very helpful, and as developers we appreciated the concomitant decrease in the number of user directed sub-searches.

Disadvantages of Hypertext: [11]

A. Disorientation: The potential for creating such a complex path that you no longer are sure how to navigate from where you are in the hypertext space to where you want to go.

B. Cognitive Overhead: The additional effort and concentration necessary to maintain several tasks or trails at one time.

C. One solution is to provide the user with screen related clues which allow them to understand how they travelled to where they are and how to return to their original focus. This limits disorientation. A second solution is to provide a limited number of levels that you will allow a user to travel before having to shift their primary focus. We accomplish this by allowing the user to browse only one level away from their main concept in the aforementioned example this was "Cellulitis". However we allow the user to perform this step an arbitrarily large number of times. For example you could run many searches for modifiers and qualifiers of Cellulitis (either at once or over multiple queries), after the first, the additional queries for modifiers and qualifiers would overwrite the preceding query. This would keep the focus on the central concept "Cellulitis" and would only loose the earlier queries for modifiers and qualifiers. Thereby maintaining orientation with a minimum of cognitive overhead.

Conclusion

In this manuscript we defined the different types of composition. Using these definitions we presented methods of both server based and client based composition. We presented many of the pitfalls involved in the creation of composition. We suggest mechanisms for minimizing these pitfalls. We have presented a simple representational schema for compound concepts involving qualifiers (concepts which change the meaning of another concept in a temporal or administrative sense) and other terms. We submit that the interaction between modifiers (concepts which change the meaning of another concept in a clinical sense) are much more complicated. One mechanism for handling this complication is to involve the user in the decision making process (User-Directed Compositionality).

Usability trial data have shown us that at present user directed composition is a mandatory component of a functional vocabulary server system. Much work is required to make server side composition a reality. User's need simple straightforward interfaces, which make clear the acceptable mechanisms for user directed composition. Such tools become a powerful mechanism for the recording of granular problems within a health information system. Optimizing for access to common modifiers and qualifiers is useful in minimizing the number of subsearches necessary to build an optimal compositional problem.

We believe it is an unconscionable pitfall to attempt to solve this problem by pre-coordinating all the combination concepts that a clinician would wish to record. This leads to an ever-expanding list of, a potentially infinite number of, compound concepts. This in turn would lead to an intolerable maintenance problem, which would surely cripple any serious long-term use of this type of vocabulary effort. Future research should be focused on the development of an ontology of modifiers. This could be similar to the ontology of qualifiers which we are currently making use of. This ontology should be constructed to contain rules regarding the allowable interactions between classes of modifiers and other types of terms, and between modifiers and qualifiers, and between modifiers themselves. By forging a better understanding of the way that modifiers interact with other terms, we can improve our post-coordinated terms. The better the system can become at anticipating results, which our users find acceptable, the greater will be our overall level of acceptance.

Another important research project would be to run the 150,000 concepts in the UMLS Metathesaurus through the Mayo-LTI vocabulary server and for each concept that has more than one match, through human review, determine if there exists a set of compound CUIs which can completely represent the base concept. For all such concepts record these relationships so that whether the user enters the separate atomic concepts or a precoordinated concept the system would know that they have the same overall meaning. An example where this is particularly important is for billing. The ICD9-CM code for a "Distal Radial Fracture" is quite different that the two ICD9-CM codes for "Fracture", "Distal Radius." This would be easily solved with the above strategy which would have on record the equivalence between the compound CUI for "Fracture"."Distal Radius" and the single concept "Distal Radial Fracture," thereby making the association between the compound CUI and the appropriate ICD9-CM code.

Compositionality is an important and necessary part of a functional controlled vocabulary system. Completeness is forever a moving target, likely never to be reached. We need mechanisms for recording clinical details at the point of care, capturing the level of granularity, which clinicians are interested in recording regarding their patient encounters. Here we present some basic definitions and our initial effort toward presenting a workable solution for a compositional model. We are currently testing this implementation in our clinical arena. However there is still much work to be done, before this complex and essential problem has been put to rest. We have suggested two additional projects, which we feel, are reasonable next steps toward the broader solution, the Complete Compositional Model.

References

[1] Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a Large Controlled Medical Terminology. JAMIA 1994;1(1):35-50.

[2] Elkin PL, Chute CG, et al. "Standardized Problem List Generation, Utilizing the Mayo Canonical Vocabulary Embedded within the Unified Medical Language System." JAMIA Symp. Suppl., in press.

[3] Rassinoux AM, Miller RA, Baud R, Scherrer JR. Modeling Just the Important and Relevant Concepts in Medicine for Medical Language Understanding: A Survey of the Issues. In Chute C (ed) Proceedings of the IMIA WG-6, 1997, 53-68.

[4] Rector AL, Nowlan WA. The Galen Project. Computer Methods and Programs in Biomedicine, 1994;45(1-2):75-8.

[5] Evans DA, Cimino JJ, Hersh WR, Huff SM, Bell DS, for the Canon Group. Toward a Medical-Concept Representation Language. JAMIA 1994;1:207-217.

[6] Elkin PL, Cimino JJ, Lowe HJ, Aronow DB, Payne TH, Pincetl PS, Barnett GO; "Mapping to MeSH"; Presented to, and Published in the IEEE Proceedings of the 12th Annual Symposium on Computers and Medical Care.

[7] Cimino JJ, Mallon LJ, Barnett GO; "Automated Extraction of Medical Knowledge from Medline Citations"; Presented to, and Published in the IEEE Proceedings of the 12th Annual Symposium on Computers and Medical Care.

[8] Baud R, Rassinoux A, Scherrer J. Natural Language Processing and Semantically Representation of Medical Texts. Meth of Inf Med 1992;31(2):117-25.

[9] Baud R, Lovis C, Rassinoux AM, Scherrer JR. Alternate Ways of Knowledge Collection, Indexing and Robust Language Retrieval. In Chute C(ed) Proceedings of the IMIA WG-6, 1997, 81-93.

[10] Chute CG, Elkin PL. Reduction of Qualifiers from a Clinically Derived Lexicon. JAMIA Sympo. Suppl., in press.

[11] Cimino JJ, Elkin PL, Barnett GO; "As We May Think: The Concept Space and Medical Hypertext"; Comput Biomed Research 1992(June 25):238-63.

[12] Weir C, Lincoln MJ, Green J. Usability Testing as Evaluation: Development of a Tool. In Cimino JJ (ed), JAMIA 1996;Symp. Suppl:870.

[13] Kushniruk A, Patel V, Cimino JJ, Barrows R. Cognitive Evaluation of the User Interface and Vocabulary of an Outpatient Information System. In Cimino JJ (ed) JAMIA 1996;Symp. Suppl.:22-26.

[14] Tuttle M, et al, Terminology Navigation in a Large Clinical Enterprise Environment, In Chute CG (ed.), Proceedings of the IMIA WG-6, 1997, 69-70.

[15] Tuttle MS, Chute CG, et al. Adding your terms and relationships to the UMLS Metathesaurus. Proc Ann Symp on Comput Applic in Med Care. 1991:219-23.

[16] Yang Y, Chute CG. A schematic analysis of the Unified Medical Language System. Proc Ann Symp on Comput Applic in Med Care. 1991:204-8.

[17] McCray AT, Sponsler JL, Brylawski B, Browne AC. The Role of Lexical Knowledge in Biomedical Text Understanding. In Stead W (ed), Proc 11th Ann Symp Comput Applic in Med Care 1987:103-107.

[18] McCray AT, Srinivasan S, Browne AC. Lexical Methods for Managing Variation in Biomedical Terminologies. In Ozbolt JG (ed), Proc 11th Ann Symp Comput Applic in Med Care, 1994:193-201.

ı Symp
l.
Capletal
fication
r appli-
Biomed

, Lavril
ELIOS
Comput
Suppl.:

lers L.
l EDI.
319-30.
ın FC,
ograms

emmel
tion of
. Meth

Iealth-
ıd Ob-
Jniver-

ı Med

Methods of Information in Medicine
© F. K. Schattauer Verlagsgesellschaft mbH (1998)

Exhibit H

J. Ingenerf[1], W. Giere[2]

[1] GSF - National Research Center for Environment and Health, Neuherberg,
[2] J. W. Goethe - University Frankfurt/Main, Germany

Concept-oriented Standardization and Statistics-oriented Classification: Continuing the Classification versus Nomenclature Controversy

Abstract: Nowadays, most activities on controlled medical vocabulari focus on the provision of a sufficient atomic-level granularity for repr senting clinical data. Amongst others, clinical vocabularies should l concept oriented, compositional and should also reject "Not Elsewhe Classified" [1]. We strongly share the opinion that there is a need to de with serious deficits of existing manually created vocabularies and with ne demands for computer-based advanced processing and exchange medical language data. However, we do not share the opinion th methodological requirements like observational and structural comparab ity needed for sound statistics should not be included in desiderata controlled medical vocabularies. Statistical-oriented classifications are n developed for representing detailed clinical data but for providing purpos dependent classes where cases of interest are assigned uniquely. Eith statistical classifications are not included into the set of controlled medic vocabularies in the sense of Cimino, or his desiderata are misleading. W argue that statistical classifications should be linked to (formal) conce systems, but again this linkage does not change their different natures. Wi this article we continue the "classification versus nomenclature" contr versy referring to Coté [2].

Keywords: Medical Informatics, Controlled Vocabulary, Documentatio Data Interpretation/Statistical

1. Introduction

Nomenclatures and classifications have been intensively studied in the fields of medical terminology, medical linguistics, medical knowledge processing, medical records and medical statistics. Standardization and classification of medical language data are dependent on the use of controlled medical vocabularies. Many of those have been developed for different purposes showing different properties. Cimino [1] lists desirable properties: domain completeness, nonredundancy, (support for) synonymy, nonvagueness, nonambiguity, multiple classification, consistency of views and explicit relationships. Especially, the property "domain completeness" is of great interest, when asking for the practical usefulness. Major vocabularies have been evaluated for their content coverage based on randomly selected medical language data out of medical records [3]. Those vocabularies that we call statistical classifications have less expressiveness and coverage (e.g. ICD, CPT) and those that we call concept systems have greater expressiveness and coverage (e.g. SNOMED). This is something we would expect because classifications inherently lead to an information loss. It should be mentioned that most existing vocabularies are variants of the vocabulary types, we are discussing in principle, or they are somehow mixtures of both (e.g., SNOMED's disease axis, that will be discussed later) [4].

The reader might argue that it is a well-known fact that old-fashioned statistical classifications should be r placed by combinatorial nomenclatur or formal compositional concept r presentation approaches. We do n agree. They should be improved t linking classes to concept represent tions in order to explicate their semar tics. But that does not mean, that the different status as statistical classific tions disappears. There is a need f vocabularies (concept systems) provi ing a sufficient atomic-level granulari for representing clinical data, but the is also a need for vocabularies (statist cal classifications) providing purpose dependent predefined classes wher cases of interest have to be assigned t uniquely. Statistical classifications ca be interpreted as views on concep systems that cannot be solely define using conceptual knowledge. We agre with Chute and Coté [5]: "The discipline of medical classification and natural language processing arise from different traditions and their application to medical patient data has been seen by some as conflicting. However, the development of semi-automated tools for storing and organizing the medical knowledge within medical patient records requires the synergistic alliance of these sciences."

In this paper we deal with the following questions:

- Why a division of the act of documentation of medical language data into "recording", "standardization", and "classification" is necessary.
- What are concept systems and statistical classifications with respect to their purposes, to their inherent ordering principles, and to additional characteristics like the computer-supported processing of them?
- What are the methodological deficits of currently available concept systems and statistical classifications, and what are the benefits when linking statistical classifications to (formal) concept systems? What are potential impacts of such a linkage for the improvement of coding software?
- Is there a meta-terminology in the field of Medical Terminology avoiding the unfortunate misunderstandings? Definitions of the terms used in this paper are provided in the Appendix.

2. Data Recording, Standardization, and Classification

As said in the abstract of this paper, current studies focus mainly on the domain-completeness and expressiveness of controlled medical vocabularies. From our point of view that is important, but generally the studies compare vocabularies that are not comparable in principle. The differences can be explained on two levels that are mutually dependent. We distinguish: (1) the level of data recording on the one hand, and (2) the level of standardization and classification on the other hand.

2.1 Answering My Own Questions Creatively versus Answering External Questions Reproducibly

(1) In general, data represent observed facts, obtained from real-world objects. Most of the time, the physician who treats the patient knows the difference between irrelevant and missing data. However, relevance is subjective. Data are also used by persons other than the responsible physician and, therefore, this difference should be made explicit, because physicians other than the treating physician do not know the reason for the absence of the data [6].

i) If no purposes for using the recorded data are known in advance the physician may enter data using free text. → No questions guide the data entry (implicit goals of observation).

ii) If purposes like coupling with decision-support systems are known in advance, entering relevant data is prescribed by explicitly asking for at least the attributes of interest for improving the completeness [7]. → Questions guide data entry (weak explicit goals of observation).

iii) If purposes like statistical evaluations are known in advance entering relevant data is guided by the rules and use of definitions of the classes of a statistical classification to select a class uniquely. At the extreme in controlled clinical trials the data entry has to conform to a study protocol [7]. → Questions and instructions for answering them uniquely guide the data entry (strong explicit goals of observations).

Generally, in practice there is a mixture of these types of data entry.

(2) Especially when we speak about qualitative medical language data, there is a need for standardizing the language data in two senses of the word:

(i) Standardizing the data by transforming the meaning into a language-independent representation based on concept systems that support the processing and exchange of that data. This kind of standardization is concept oriented and carried out by what is called indexing [8, 9]. Amongst others, the recognition synonymous or hyponymous e pressions is supported. It is primar ly oriented to the individual. Th term "standardization" should b reserved exclusively for this inte pretation, as it will be done in th paper. The vocabularies needed fo standardization are called "concep systems" in this paper. Combinato rial multi-hierarchical vocabularie like the SNOMED-nomenclatur and MeSH-thesaurus are non-formal, manually created variants o concept systems. The desiderat from Cimino [1] are applicable to this kind of vocabularies. We avoid the ambiguous term "classification" for this kinds of taxonomically structured concept systems.

→ A concept system provides a "lexicon" and a "grammar" for representing facts (What can be said?). It does not provide questions guiding the data entry (What should be said?).

(ii) Physicians entering free text can take notice of weak explicit goals of observation. Alternatively, structured data entry [10] or predictive data entry [11] leads to more complete and reliable data. In both cases standardization should be performed similar to (i).

(iii) Medical language data is "standardized" based on explicit goals of observation with respect to a question, guiding the collection of data uniformly. From that, differentiation criteria and value-categories for data entry are derived, allowing data ordering (register). This interpretation of the term "standardization" is used by most people (and also in this paper) when speaking about "classification" of data. It is primarily population oriented. Intra- and interobserver variability are reduced as much as possible. The vocabularies needed for the classification, especially of complex items like diagnoses, are called "statistical classifications" in this paper. The desiderata from Cimino [1] are not applicable to these kinds of vocabularies.

→ A statistical classification prescribes questions and instructions for answering them (What should be said?). It does not provide the detail for representing facts semantically.

Amongst others, different interpretations of the terms "standardization" and "classification" are probably responsible for the mentioned controversy in the title of this paper. If there are no explicit goals of observation in advance it is of course important to provide detailed standardized data based on expressive vocabularies in order to re-classify as good as possible for purposes that arise in the future. However, as the data have not been collected with the goals of potential retrospective studies in mind, the values for some variables of interest typically are unavailable, and the sample population may not represent the ideal study population. Typical examples of biases are illustrated in [12]. Independent of that "prospective versus retrospective" issue (level 1) the different status of statistical classifications as vocabularies (level 2) remains true. The three types of data entry described above should be used in combination as illustrated by the slogan: "Without explicit goals of observation no scientific knowledge, no scientific knowledge just with explicit goals of observation [13].

Example: We refer to the TNM-classification [14]. Imagine there is a patient with a larynx carcinoma that infiltrates the supraglottis. Without going into detail, think about possible observed facts that might be relevant for characterizing this case. These facts are written down as complete as possible and standardized as detailed as possible based on a concept system. Now, the patient is no longer available and you try to re-classify the case according to the TNM-classification using only the textual representation. Can you be sure, your description includes that more than one part of the supraglottis is infiltrated and that the vocal cord is movable (needed for assigning to T2)? Can you derive the information that several ipsilateral lymph nodes of the neck are attacked and that not one of them is enlarged greater than 6 cm (needed for assigning to N2b)? Is the information about histopathological grading for the primary tumor available, and when 'yes' is the assignment to the classes "well good differentiated" (G1) up to "undifferentiated" (G4) or "not judicable" (GX) possible? Much more information is needed to complete the organ-specific classification into the TNM-system regarding the detailed instructions to guarantee the selection of one and only one class. The differences between the classes are very subtle and have serious consequences for prognosis and therapy. If the meaning of the classes of the TNM-classification are explicated by detailed concept representations and if the textual representation of the case is transformed also in a concept representation it is possible to support the mapping to some extent automatically. And we agree that such an approach can improve the effectiveness and correctness of the classification of the case (see last section of this paper). However, we cannot believe that anybody is convinced that such a mapping can be provided fully automatically with all the consequences for the patient. And if yes, again that does not mean that the TNM-classification is just some kind of a concept system with less granularity. It is a statistical classification, "asking" you for attributes and predefined values that are necessary for its purpose and giving instructions for answering the "questions" uniquely.

2.2 *Data Collection versus Data Ordering*

What has been outlined in the previous section can be best summarized coherently by Fig. 1. It shows a modified version of the BAIK information model from Giere [15], that has guided his work for more than 20 years.

In the first two care-oriented cycle: patient comes with a problem (?) to physician. He examines the patient ai notes symptoms, signs and tests in tl medical record depending on his exp rience. The medical record gives hi the individual information he needs treat (!) the patient.

Mainly for the purpose of inform tion retrieval, free text from the medic record is indexed, such that it is sta: dardized with respect to medical la guage phenomena leading to compa ible information. Alternatively, tl structured data entry approaches con bine the second and third cycle in th: free text narratives are minimized (se points (i) and (ii) in the previou section). Passing this third cycle ca support significantly the creation c registers in the fourth cycle. Mainly fc the purpose of statistical analyses, dat from the medical record are selected t be classified with respect to predefine differentiation criteria depending o the question(s) of interest (see poin (iii) in the previous section). This lead to comparative information. Both kind of information are added to the individ ual information of a single patient. The are typically used for teaching, qualit assurance, support of administrativ management and clinical research.

From the register, statistical informa tion can be drawn which allows the re searchers to formulate a hypothesis (?) This can be verified or falsified in a statistical experiment, i.e. a controlled statistical study (see point (iii) in the previou section). The resulting knowledge (!) again adds general information to the comparative, compatible and individual one.

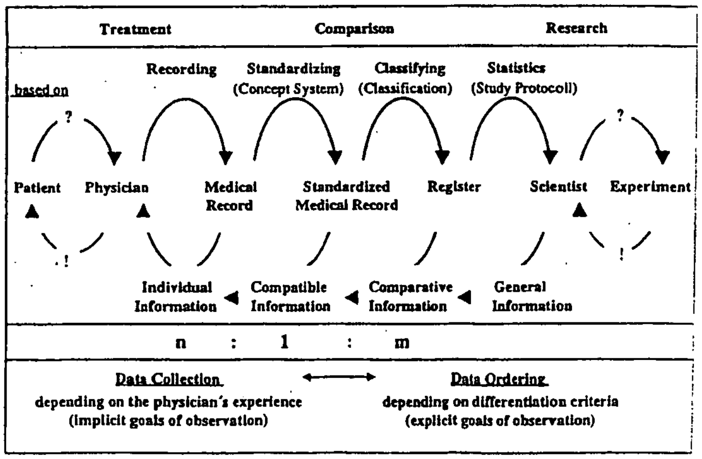


Fig. 1 Data collecting and data ordering.

Table 1 Comparing data collection and data ordering.

Data Collection	Data Ordering
Patient	Case
Individual	Population
Characterizing	Typing
Identification	Discrimination
Communicative	Distributive
Implicit goals of observation	Explicit goal of observation
Open for new terms	Closed and predefined
Creative	Reproductive
Lifelong	Episode
Primary	Secondary

It should be mentioned that in the transition from individual to compatible information, from compatible to comparative information, and from comparative to general information there is in each step a tradeoff between loss of information on the one hand, and getting an added value on the other hand. The added value of standardization can be explained by the fact that several (n) individual information units describing one patient and collected by different physicians possibly in different countries can be transformed into compatible information units and, therefore, integrated into one (1) coherent language-independent representation ready for exchange. Given the restrictions explained in the previous section it is possible to classify this standardized information unit according classifications with respect to different purposes (m). Overall the difference between data collection and data order is highlighted in Table 1 [15]:

3. Concept Systems and Statistical Classifications

The distinction between standardization and classification as activities described in the last chapter is now transferred to the distinction between concept systems and statistical classifications as schemes in this section.

3.1 The Semiotic Triangle and Ordering Principles

Existing controlled vocabularies are generally based on ordering principles. These principles can be described with respect to several corners in the semiotic triangle, depending on the purpose of the vocabulary. Figure 2 shows an extended version of this triangle. With respect to Wüster [16], topics ar included that indirectly refer to concrete documents in which they ar mentioned.

Referring back to Fig. 1 and goin from right to left, general informatio (knowledge) is stored in electroni libraries. Suppliers like the Nationa Library of Medicine provide content preserving and foreseeable selections o keywords from a thesaurus like MeSH [17] for both, the indexer and the searcher. The ordering principle of a thesaurus is focused on topics in Fig. 2. This will not be discussed in this paper (see [18]).

The generation of valid comparative information, as outlined in Fig. 1, is dependent, amongst others, on clear instructions for inclusion and exclusion objects in classes, designated by rubrics. Hence, statistical classifications like ICD are used in order to ensure unique assignments of objects to one class. Otherwise class frequencies cannot be interpreted statistically. The ordering principle is primarily focused on the concept's extension enabling a disjoint partition of a set of (recorded) objects. More correctly, the semiotic triangle should be complemented by an analogous "rubric" triangle in Fig. 3, covering also rubrics and classes.

Compatible data, as presented in Fig. 1, can be exploited very well by information-retrieval techniques and can be augmented by data-driven techniques for decision support, quality assurance, and instruction. This is because the processing of standardized data can exploit all the detail and inherited concept relations and attributes provided by concept systems. They are primarily organized with respect to the concept's intension. In nomenclatures like SNOMED also the terms in Fig. 2 are taken into consideration systematically.

Finally, individual information as presented in Fig. 1 is provided and interpreted by a physician depending on his skills and interests, previous findings and others. Considering the properties "creative" and "open for new terms" in Table 1, the pragmatic context of the physician is not covered by present vocabularies properly but should be taken into consideration. A medical report from a younger

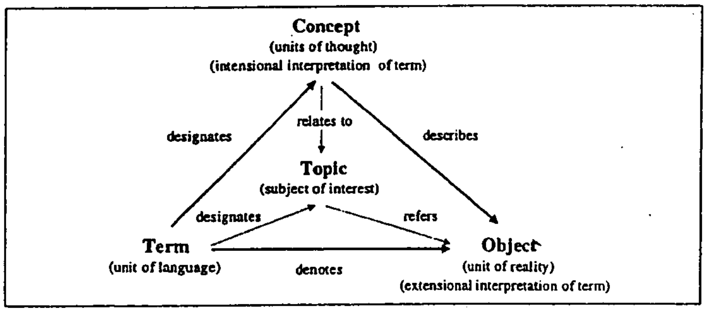


Fig. 2 Extended semiotic triangle.

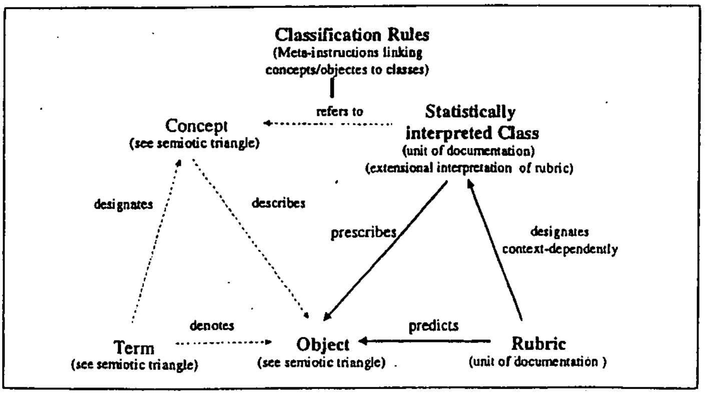
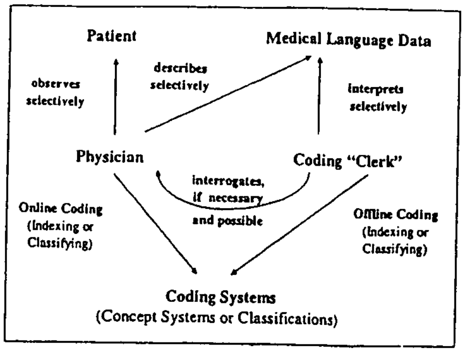


Fig. 3 Rubric triangle.

colleague has a different pragmatics than one from an experienced colleague. Findings reported in a university hospital have another pragmatics than those reported in primary care. It is often the individual's creativeness and intuition outside of an established scheme or vocabulary that leads to new scientific knowledge (see also [19]).

3.2 *What can be said? versus What ought to be said?*

Concept systems and statistical classifications usually use codes for the unique identification of concepts and rubrics. Therefore, they are uniformly called coding systems. The mapping of medical language data into coding systems is called coding, independent of differences between indexing and classifying. Fig. 4 illustrates two levels of coding data that should be distinguished.



Patient
Medical Language Data
observes selectively
describes selectively
interprets selectively
Physician
Coding "Clerk"
interrogates, if necessary and possible
Online Coding (Indexing or Classifying)
Offline Coding (Indexing or Classifying)
Coding Systems
(Concept Systems or Classifications)

Fig. 4 Online versus offline coding.

Online coding has advantages in that the physician usually has greater knowledge about the patient and also about the subject field. A correct classification often needs judgements that can be provided solely by the physician. A disadvantage is that he needs good knowledge and experience in the use of coding systems. Offline coding has advantages in that the coding clerks usually have more experience in the use of coding systems. Especially for indexing, the effort for coding can be shifted to a computerized "coding clerk". This is possible because concept systems provide a conceptual lexicon and grammar suitable for natural language processing approaches [8, 9]. However, as outlined earlier, concept systems do not provide explicit questions and instructions for answering them uniquely. They offer concepts for standardizing medical language data in a passive manner answering the question "What can be said?".

Statistical classifications do provide questions and instructions for answering them. Usually they are reflected implicitly by the classification's structure. They ask top-down for division criteria establishing disjoint sub-classes in a goal-oriented, active manner. There are meta-statements like notes, exclusions and inclusions for guiding the selection and the intended interpretation of rubrics. As illustrated in Fig. 3, classification rules help a coder to correlate the rubric with what is actually written in a document. Also, the context of rubrics can direct the selection of a proper class, mainly based on negative decisions, i.e. "Not Elsewhere Classified". Generally, these so-called residual classes have the code extension ".8". They are needed to guarantee the completeness feature of statistical classification. The meaning of residual classes can be derived by the complement of the meaning of all sibling classes with respect to the superior class. Finally, subclasses are offered where the division criterion is unknown, i.e. "Not Otherwise Specified". Generally, these classes have the code extension ".9". In mono-hierarchical classifications, like the ICD, this is a prerequisite for getting complete information with respect to the underlying goal, in that coding up to a defined digit is prescribed. The reasons for the absence of the needed information can be refined in order to distinguish (amongst others) between missing, irrelevant, and not obtainable data.

As a classification of data leads to an information loss, the consequence is that already classified data according to one classification cannot be re-classified according to another. Two classifications are generally incompatible with respect to each other. That is also true with different versions of one classification, e.g. in the 9th and 10th version of the ICD [20]. On the other hand, standardized data based on concept systems offer a greater flexibility and openness for unforeseeable uses [21]. The assumption is made, that data is represented as detailed as possible with enough discriminatory power for new purposes. But there appear to be other purposes where the detail is unnecessary and where, ironically, needed data are missing. Completeness of data with respect to the physician's view is different from completeness of data with respect to an external purpose-dependent view. Haux has presented a conflict between data recording guided by decision support component on the one hand and appropriate statistical data analysis on the other hand due to selection biases [22].

3.3 *Statistical Classifications = Concept Systems + Explicit Instructions for Recording Reproducibly?*

The reader might argue that statistical classifications can be redefined as concept systems together with explicit stated instructions for recording reproducibly. This equation makes sense from our point of view and it explains somehow the relationship between these two types of controlled vocabularies. However, that does not mean that statistical classifications as a separate type of vocabulary can be substituted by concept systems as the only acceptable type of vocabulary fulfilling the desiderata of Cimino [1]. There are two reasons: First, although the meaning of statistically motivated classes can and should be explicated to some extent by linking them to concept representations, this reduction cannot be provided completely within the "world" of concept systems. Reasons are:

- Mutual exclusiveness of classes as well as residual classes must be defined by extensive use of the logical operators "negation" and/or "disjunction" for enumerating concepts that are included or excluded in classes. This contradicts with practical or fundamental limits of what can be expressed with concept systems, especially when using formal approaches like description logics or conceptual graphs.
- Meta-instructions like preference rules for criteria (e.g. etiology prior to manifestation) or notes for human interpretation of classes (e.g. for psychological diseases) as well as references to the act of documentation (e.g. in "not otherwise

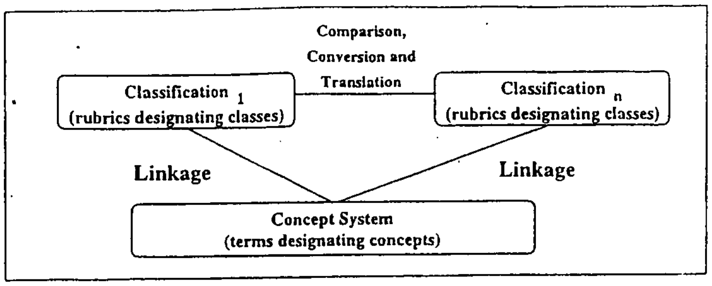

Comparison, Conversion and Translation
Classification 1 (rubrics designating classes)
Classification n (rubrics designating classes)
Linkage
Linkage
Concept System (terms designating concepts)

Fig. 7 Classification linked to concept systems.

- Language- and purpose-independence as far as possible:
  - Multilingual and flexible generation of free text.
  - Sharing and re-use of terminology is supported as requested views on concepts can be offered, i.e. mappings into existing coding systems.
- All terminological services can be offered to computer applications for "any" composite concepts of interest.

With such an approach it will be possible to resolve most of the deficits of SNOMED mentioned above. For example, in SNOMED the following doubtful concept relation is included: M-8895 "Myoma" subordinated under M-889 "Leiomyoma". Given the following concept definitions:

Myoma ≐ Tumor which has_Location Muscle Tissue
Leiomyoma = Tumor which has_Location Smooth Muscle Tissue
Smooth_Muscle_Tissue = Muscle_Tissue which has_Type Smooth

Then the following generic relations can be deduced, just regarding the formal definitions and the inheritance of characteristics by the so-called classifier:

SmoothMuscleTissue < MuscleTissue
Leiomyoma < Myoma

The deduced generic relation between both concepts is just inverse to the one offered in SNOMED. Given a huge number of concepts defined in this way, all the implicit knowledge hidden in a concept system can be made explicit. With that, the consistency and non-redundancy of a concept system can be ensured effectively. This again is a good pre-requisite for maintaining, updating, and cross-referencing with other coding schemes, etc. Finally, the implementation of concept systems as a terminology server (like the one in GALEN) can effectively provide terminological services to other computer applications, like data entry systems, and knowledge and language processing systems.

All the advantages are not without concerns. The used formal concept representation languages are controversially discussed in the knowledge representation community with respect to a trade-off between expressive power and the computational costs, or even intractability of the language [29]. Approaches with high expressive power and reasonable computational costs generally make concessions to soundness and completeness of the implemented algorithms. Ceusters et al. have some doubts concerning two of the basic assumptions with formal concept systems: the purpose- and language independence [30]. These doubts, together with additional questions like "Can a formal concept system be built that can be scaled up to realistic systems?" are dealt with seriously, especially in the GALEN project [26]. For many special issues like dualities (e.g. "The erosion lasted for three month" is interpreted as process or lesion?), paradoxes with conjunction and regions (e.g. "Ulceration of the GI tract from stomach to duodenum" is located where?) etc. there are still no complete answers available [31].

4.4 Deficits and Improvements of Statistical Classifications

Concept systems in the sense of systematic nomenclatures, like SNOMED, are in general regarded as too complicated to be handled manually. Furthermore, formal concept systems can only be used effectively by computerized implementations of terminology servers. Contrarily, classifications are somehow easier to handle. Hirs gives a reason for this [32]: "The reason why classification (as an activity) can be treated as a separate stage of human information processing is its restriction of the number of relations into one direction: the whole. There is a generic 1:N relation between genus and species. The generic relation has the advantage that a classification can cope with complex relationships between objects, characteristics, and concepts, e.g. by assigning objects (diseases, medical procedures etc.), that can be counted. However, the generic relation generates at the same time some semantic problems of human information processing. Objects have to be considered within a specific field, a given point of view and a fixed sequence of subdividing characteristics. This is why the ICD, like (almost) every classification, is a compromise between conflicting interests."

Figure 6 illustrates most of the specific characteristics of statistical classifications that have been discussed earlier: the mono-hierarchy, the mutual exclusiveness and completeness of subclasses with respect to the superior class, the use of residual classes (e.g. 320.8 can and should be selected, if hemophilus, pneumococcal, streptococcal and staphylococcal meningitis can be excluded), and the use of meta-instructions like notes (e.g. 326) and meta-constructions like the "Not Otherwise Specified" – classes (e.g., 320.9).

4.5 Deficits of Existing Classifications

Most of the mentioned deficits of existing concept systems in the last section apply also for classifications, i.e., the combinatorial explosion, redundancy and inconsistency, implicit and context-dependent semantics. Most of the deficits are necessary consequences of the underlying ordering principle of statistical classifications. There was a basic strategy for dealing with these deficits when speaking about concept systems in the last section: removal of

Original rubric and intermediate representation	Formal representation inferred from the intermediate one
RUBRIC "dividing of papillary muscle" CODE "xy" MAIN deed:dividing ACTS_ON anatomy:papillary muscle HAS_OTHER_FEATURE method VALUE induced arrest of heart	(SurgicalDeed which isCharacterizedBy (performance whichG isEnactmentOf (Dividing whichG < playsClinicalRole SurgicalRole actsSpecificallyOn PapillaryMuscle hasSubprocess InducedCardiacArrest>))) The left intermediate representation is somehow similar to the way of combinatorial representation with SNOMED.

Fig. 8 Linking a procedure classification to the GALEN concept system.

complex pre-coordinated concepts by defining them based on simpler ones when needed. A solution for this problem was, at the same time, one for most of the other problems. This basic strategy for dealing with deficits of statistical vocabularies is not available. Monohierarchical classifications are not generative, at least not in this way. The rubrics (classes) cannot be substituted by combinations of subrubrics (subclasses), amongst others, due to their complex context-dependent semantics. One deficit of statistical classification that is most often criticized is the poor expressiveness and domain-completeness in order to represent clinical data as detailed as possible. From our point of view, this is not a deficit but a necessary property.

4.6 Improvement of Existing Classifications

The only way to deal with the deficits of classifications is to take rubrics as they are, and to make explicit as much as possible their contextual meaning with respect to attached concept systems (Fig. 7).

The development, test of integrity and redundancy, and the maintenance of one classification can be supported by linking each class to a concept representation in order to make explicit its meaning. Furthermore, the comparison, conversion and translation of classifications can be carried out most effectively by linking them to concept systems instead of linking two classifications directly. This has been done very early, for example, by Thurmayr based on SNOMED [33]. The criteria used for subdivision in classifications can be made explicit for an integrity test. The development of conversion tables between two classifications can be supported. The same idea is used in an initiative of the European Federation of Classification Centers (EFCC). In cooperation with the GALEN project, national procedure classifications are studied and harmonized using appropriate tools based on the terminology services described earlier. It should be mentioned, that this collaborative work on harmonizing national classifications relies on standardized meta-models for studying rubrics of existing coding systems and describing medical procedures semi-formally and coherently. Such meta-models are developed in the working group II on Surgical Procedure Modeling of the CEN TC251. Figure 8 illustrates an example. The whole transformation process and the support by tools providing specific terminological services is described in [34].

Finally, SNOMED itself offers an example of this idea. The disease and procedure axes are actually classifications next to the rest of SNOMED. Typically, pre-coordinated and complex concepts like syndromes designated by eponyms like "Cohen syndrome" are provided, if possible with links to the other axis. Within SNOMED, the relation between a nomenclature and a classification is explained with the example: "Tuberculosis (D) = Lung (T) + Granulom (M) + M.Tuberculosis (E) + Fever (F)" [35]. Furthermore, cross-references to the ICD-9-CM classification are added. However, the properties of a statistical classification (see the Appendix) are not fulfilled by the disease axis. Concepts at one level are not mutually exclusive and frequently the residual and "explicitly stated as unknown"-classes are omitted, but needed for exhaustiveness. Finally, the fourth property in the definition, is not reflected in the disease axis. Subconcepts are listed for terminological reasons, not taking into consideration the expected frequencies concerning the concepts' instances. It seems that the disease axis can be considered more as a concept-oriented alphabetical index for the ICD-9-CM.

4.7 Deficits and Improvements of Coding Software

Manual coding is still a burdensome task with a significant error rate. Accurate coding requires the coders to have a detailed knowledge of the coding system. The resulting codes are used for statistical analysis, but not for patient treatment. There is a motivation problem. Interactive coding software can improve the situation. However, existing approaches are just able to apply string searching for comparing the input text with the rubrics of the underlying classification. Together with other functionality, like navigating through the hierarchies, this is highly interactive and time consuming. Also, string searching is not powerful enough to guarantee that all relevant parts of the classification are regarded.

Referring to Fig. 1, instead of just classifying the original individual data and skipping the third cycle, it is advantageous to classify based on already standardized data. Now, there are several approaches supporting this idea. Some of the approaches are already realized in computerized coding systems. Very conventionally, the alphabetical index of a classification is already a means for linking medical descriptions with classes of a classification. This can be extended by the provision of a huge set of pre-coded medical phrases from everyday language [36]. Yet another approach is the provision of user-defined and department-specific micro-glossaries [37].

When linkages of the entered medical phrase and of the classification to concept systems are established a function similar to the alphabetical index can be exploited. Instead of comparing two classifications, an entered medical phrase can be mapped to rubrics of a classification more effectively. Classifying can now be based on comparing conceptual representations instead of strings (Fig. 9).

Wingert et al. [9] and Satomura and Do Amaral [38] established transformation tables between SNOMED-code combinations and ICD-codes. At the same time they provided language pro-

Table 3 Indexing and classifying.

Indexing	Classifying
Precise semantic characterizing with concept representation	Unique assignment to one class
Selecting and composition of relevant concepts	Assignment to a class with respect to other classes
Predicative interpretation of a ∈ Concept	Membership interpretation of a ∈ Class
Completeness and relevance of data with respect to the physicians view	Completeness and relevance of data with respect to a purpose dependent view
Automated indexing is feasible	At best semi-automated classifying is feasible, needing feedback-interaction

cessing systems for indexing medical language data based on SNOMED automatically. From that, a mapping to ICD is continued formally. It should be mentioned that the experience showed that the frequency of mapping into ".9"-classes (unspecified-classes) is much higher compared to an interactive classification. That is obvious after having read the second section of this paper. It should be possible to complete missing data by feedback questions to the physician for improving data quality.

Similar to the linkage of ICD-classes to a concept system, Delamarre et al. have linked classes of ICD-9-CM to conceptual graph representations making the rubric's context as explicit as possible [39]. The incoming diagnoses are translated to conceptual graphs, as well using a language-processing system within the European MENELAS project. From that, a mapping between is carried out by comparing conceptual graphs. As already mentioned, the conceptual reconstruction of classes can support but not substitute a classification. As most of the mentioned approaches start from already written medical text, i.e. they are offline-coding approaches regarding Fig. 4. Necessary data are likely missed, therefore such a mapping should be at least interactive. Table 3 summarizes the differences between indexing and classifying.

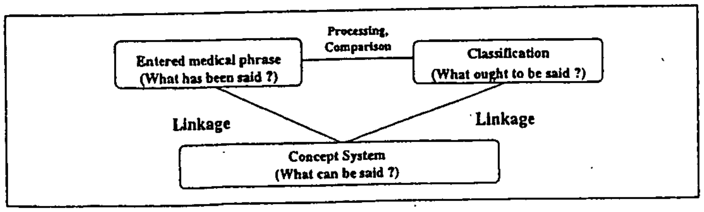


Fig. 9 Medical phrase and classification linked to concept systems.

5. Conclusion

The special characteristics of and the difference between concept-oriented standardization and statistical-oriented classification should have been made clear. We think that this distinction is exactly one way to end the mentioned controversy in the title. The two interpretations of standardization and classification contribute significantly to it. At least, students learning the fundamentals of the field of medical terminology are confused when basic terms are not defined and not used uniformly.

In many controversial discussions on the topic of this paper we noticed that the mix of two different levels causes much confusion. On one level there is the question of the best way to record medical data. The advocates of an uncontrolled documentation recommend a detailed and faithful representation of clinical data with enough discriminatory power for new purposes, like re-classification for retrospective studies [40]. The advocates of a controlled documentation recommend to plan documentation in advance in order to ensure complete and reliable data for potential uses [7]. By the way, both points of view do not exclude each other. However, both points of view have special attitudes how to deal with standardization and especially with classification of medical language data. Independent of these attitudes, there is the other level that deals with concept systems and statistical classifications such as coding schemes. This issue is much more difficult because it causes many theoretical consequences. Vocabularies for coding and classification in the seventies are discussed under the umbrella of knowledge representation languages since the late eighties [1]. The roots of this process, the methodological considerations for statistically-oriented data classification seems to get lost. We do not want to go back to the roots, but propose that synergistic alliance mentioned in the quotation of Chute and Coté [5] in the introduction.

We will end this paper with a vision. The physician dictates his report and the system is able to recognize the speech, translates it into written text, and indexes it based on a concept system. New medical language is recognized. The terminological and language knowledge resources can be updated so that they reflect the used clinical language. Based on this standardized representation, several sources of patient data can be integrated properly, can be exchanged and linked to decision support systems. Given new purposes for clinical research, the system is able to recognize missing data online and to complete it by offering feedback-questions to the physician.

Acknowledgments

The authors would like to thank their colleagues at the GSF-Medis-Institute and the J. W. Goethe-University Medical School for valuable discussions. We appreciate the working groups "Medical Documentation and Classification" chaired by Rüdiger Klar and "Medical Linguistics" chaired by Peter Pietrzyk within the German Medical Informatics Society (GMDS) for valuable discussion and working forums.

Josef Ingenerf would also like to thank the GALEN-IN-USE-community for many valuable discussions and the opportunity to get into the international medical terminology "community". Finally, this paper would not have been possible without the influence and patience of my various mentors, especially Reinhold Haux, Walter Oberschelp and Rudolf Repges.

REFERENCES

1. Cimino JJ. Desiderata for Controlled Medical Vocabularies in the Twenty-First Century. Meth Inform Med 1998; 37: 394-403.
2. Coté RA. Editorial: Ending the classification versus nomenclature controversy. Medical Informatics 1983; 8: 1-4.
3. Chute CG, Cohn SP, Campbell KE et al. The content coverage of clinical classifications. Journal of the American Medical Informatics Association 1996; 3 (3): 224-33.
4. Cimino JJ. Review paper: coding systems in health care. Methods of Information in Medicine 1996; 35 (4-5): 273-84.
5. Chute CG, Coté RA. Computerized Natural Medical Language Processing for Knowledge Representation: Overview of IMIA Working Group Conference, Geneva, September 1988. In: Barber B, Cao D, Gin D, et al., eds. *Proc. of the MEDINFO 89 - 6th International Conference on Medical Informatics in Peking/Singapure.* Amsterdam: North-Holland, 1989: 878-81.

6. van Bemmel JH. Handbook of Medical Informatics. In: Houten/Diegem: Bohn Stafleu Van Loghum, 1997.
7. Leiner F, Haux R. Systematic Planning of Clinical Documentation. Methods of Information in Medicine 1996; 35: 25-34.
8. Baud RH, Rassinoux AM, Wagner JC et al. Representing Clinical Narratives Using Conceptual Graphs. Methods of Information in Medicine 1995; 34 (1/2): 176-86.
9. Wingert F, Rothwell DJ, Coté RA. Automated indexing into SNOMED and ICD. In: *Proc. of the IMIA-WG6 International Working Conference in Geneva, 1988.* Scherrer J-R, Coté RA, Mandil SH, eds. Amsterdam: North-Holland, 1989: 201-39.
10. Moorman PW, van Ginneken AM, van der Lei J, van Bemmel JH. A model for structured data entry based on explicit descriptional knowledge. Methods of Information in Medicine 1994; 33: 454-63.
11. Nowlan WA, Rector AL. Medical Knowledge Representation and Predictive Data Entry. In: *Proc. of the AIME 91 - 3rd Conference on Artificial Intelligence in Medicine Europe.* Stefanelli M, Hasman A, Fieschi M et al., eds. Berlin: Springer 1991: 105-16.
12. Feinstein AR. Scientific Standards in Epidemiologic Studies of the Menace of Daily Life. Science 1988; 242: 1257-63.
13. Giere W, Schuster RW. Informationsaustausch zwischen Krankenhaus und Praxis. Der Praktische Arzt, Arzt für Allgemeinmedizin 1975; 8: 1-4.
14. UICC. *TNM Classification of Malignant Tumours, 4th Edition (edited by P Hermanek, LH Sobin).* Berlin: Springer 1987.
15. Giere W. The BAIK Model. In: *Open Systems in Medicine.* Fleck E, ed. Amsterdam: IOS Press 1995: 24-34.
16. Wüster E. Begriffs- und Themaklassifikation: Unterschiede in ihrem Wesen und in ihrer Anwendung. Nachrichten für Dokumentation 1971; 22 (3/4): 98-104 & 143-50.
17. National Library of Medicine (NLM). *Medical Subject Headings (MeSH): Tree Structures (Thesaurus for searching in the online-database "MEDLINE").* Bethesda, MD: Nlm-Med-85-03, 1985.
18. Ingenerf J. Taxonomic vocabularies in medicine: the intention of usage determines different established structures. In: *Proc. of the MEDINFO 95 - 8th Int. Conf. on Medical Informatics in Vancouver, BC, Canada.* Greenes RA, Peterson HE, Protti DJ, eds. Amsterdam: North-Holland 1995: 136-9.
19. Frutiger P. Language and knowledge interfaces: progress towards IMIA's WG6 recommendations. In: *Proc. of the MEDINFO 86 - 5th International Conference on Medical Informatics in Washington, DC.* Salamon R, Blum BI, Jorgensen M, eds. Amsterdam: North-Holland 1986: 76-80.
20. Zaiss A, Schulz S, Graubner B, Klar R. Conversion table between ICD-9 and ICD-10. In: *Proc. of the MIE 96 - 12th International Conference on Medical Informatics in Europe in Copenhagen (Denmark).* Brender J, Christensen JP, Scherrer J-R et al., eds. Amsterdam: IOS Press 1996: 193-7.
21. Rector AL, Nowlan WA, Kay S. Foundations for an electronic medical record. Methods of Information in Medicine 1991; 30: 179-86.
22. Haux R. Knowledge-based decision support for diagnosis and therapy: on the multiple usability of patient data. Methods of Information in Medicine 1989; 28: 69-77.
23. Wingert F. Automated Mapping of ICD into SNOMED. In: Orthner HFBBI, ed. 1989: 199-204.
24. Coté RA. International Classification for Health and Disease: the expandable common core concept. Medical Informatics 1983; 8: 5-16.
25. Rothwell DJ. SNOMED based knowledge representation. Methods of Information in Medicine 1995; 34: 209-13.
26. Rector AL, Nowlan WA, Kay S. Goals for concept representation in the GALEN project. In: *Proc. of the SCAMC 93 - 17th Annual Symposium on Computer Applications in Medical Care, Washington DC.* Safran C, ed. New York: McGraw-Hill 1993: 414-8.
27. Campbell KE, Das AK, Musen MA. A logical foundation for representation of clinical data. Journal of the American Medical Informatics Association 1994; 1 (3): 218-32.
28. Ingenerf J. On the relationship between description logics and conceptual graphs - with some references to the medical domain. In: *Proc. of the 19th Annual Conference of the Gesellschaft für Klassifikation (GfKl) in Basel, Switzerland, March 1995.* Bock HH, Posasek W, eds. Berlin: Springer 1996: 355-69.
29. Doyle J, Patil RS. Two theses of knowledge representation: language restrictions, taxonomic classification, and the utility of representation services. Artificial Intelligence 1991; 48: 261-97.
30. Ceusters W, Deville G, Buekens F. The Chimera of Purpose- and Language Independent Concept Systems in Health Care. In: *Proc. of the MIE 94 - 12th International Conference on Medical Informatics in Europe, Lisbon (Portugal):* Barahona P, Veloso M, Bryant J, eds. 1994: 208-12.
31. Rector AL. Coordinating Taxonomies: Key To Re-usable Concept Representations. In: *Proc. of the AIME 95 - 5th Conference on Artificial Intelligence in Medicine Europe.* New York: Barahona P, Stefanelli M, Wyatt J, eds. Springer 1995: 17-28.
32. Hirs WM. The use of terminological principles and methods in medicine. In: *Proc. of the MEDINFO 92 - 7th International Conference on Medical Informatics in Geneva Palexpo, Switzerland.* Lun KC, Degoulet P, Piemme TE et al., eds. Amsterdam: North-Holland 1992: 1452-7.
33. Thurmayr R. Use of a Conversion Table from ICD/E-Code to SNOMED. In: *Proc. of the IFIP-IMIA WG6 International Working Conference in Ottawa, 1984.* Coté RA, Protti DJ, Scherrer J-R, eds. Amsterdam: North-Holland 1985: 333-7.
34. Rodgers FE, Solomon WD, Rector AL et al. Rubrics to Dissections to GRAIL to Classifications. In: *Proc. of the Medical Informatics Europe (MIE) '97.* Pappas C, Maglaveras N, Scherrer J-R, eds. Amsterdam: IOS Press. 1997: 241-5.
35. Coté RA, Rothwell DJ, Palotay JL et al. *Systematized Nomenclature of Human Veterinary Medicine - SNOMED Interna nal (4 volumes).* Chicago: College of Am can Pathologists 1993.
36. Kolodzig C, Diekmann F. Diagnosenkoc rung: Nutzung von DV-Systemen (wie DIACOS). Deutsches Ärzteblatt 1992; (3): 92-6.
37. Assimacopoulos A, Le Coultre C, Gries V, Scherrer J-R. Nomenclature or Classifi tion? - Eighteen Months of Interact Coding of Diagnosis and Surgical Pro dures Within the Integrated Hospi Information System (HIS) DIOGENE. *Proc. of the MEDINFO 86 - 5th Internatio Conference on Medical Informatics Washington, DC.* Salamon R, Blum Jorgensen M, eds. Amsterdam: Nor Holland 1986: 865-9.
38. Satomura Y, do Amaral MB. Automat diagnostic indexing by natural langua processing. Medical Informatics 1992; 149-63.
39. Delamarre D, Burgun A, Seka LP, Le Be P. Automated Coding of Patient Dischar Summaries Using Conceptual Grap Methods of Information in Medicine 19 34 (4): 345-51.
40. Rector AL. Faithfulness or comparabili Methods of Information in Medicine 19 35 (3): 218-9.
41. ISO 1087. *Terminology - Vocabula* Berlin: Beuth 1990.
42. ISO 5127/6. *Documentation and informati (Vocabulary), Part 6: Documentary la guages.* Berlin: Beuth 1983.
43. Rescher N. *Introduction to Logic.* Ne York: St. Martin's Press 1964.
44. World Health Organization (WHC *Manual of the International Statistic Classification of Diseases, Injuries, a Causes of Death - Ninth Revision of t International Classification of Disease* Geneva: WHO 1977.
45. Coté RA, Rothwell DJ. The classification nomenclature issues in medicine: a return natural language. Medical Informatics 198 14 (1): 25-41.
46. CEN-ENV-12264. *Medical Informatics Categorical structure of systems of co cepts - Model for representation of semantic* Brussels: CEN 1995.

Address of the authors:
Dr. Josef Ingenerf,
Medical University of Lübeck,
Institute of Medical Informatics,
Ratzeburger Allee 160,
23538 Lübeck,
Germany
E-mail: ingenerf@medinf.mu-luebeck.de

Appendix

Annotated Definitions with Examples

Object (ISO 1087 [41]): Any part of the perce vable or conceivable world.

Concept (ISO 1087 [41]): A unit of thought cor stituted through abstraction on the basis c properties common to a set of objects.

Characteristic (ISO 1087 [41]): Mental representation of a property of an object serving to form and delimit its concept.

Intension (ISO 1087 [41]): Set of characteristics which constitute a concept.

Type of Characteristic (ISO 1087 [41]): Any category of characteristics used as a criterion for the establishment of a generic system of concepts.

System of Concepts (ISO 1087 [41]): Structured set of concepts established according to the relations between them, each concept being determined by its position in this set.

Concept Systems (as used in this paper): System of concepts where the generic relation is based as much as possible on the intension of concepts.

Remark 1: Concept systems are especially characterized by partitative and generic hierarchical relations between concepts, establishing a taxonomic, subsumption, or type hierarchy. The generic relations from one concept to other concepts are induced by the Aristotelian kind of concept definitions providing genus and differentia (e.g. "Viral infections are infections caused by virus"). The definition gives deductive arguments to analyze the inheritance of characteristics and to determine the concept's position in the taxonomy according to levels of inclusiveness, also called concept classifying. Using a formal reconstruction of this idea, a computable management of concept systems is possible. It should be noted that the set (measles, mumps, herpes, etc.) represents the class or extension of the concept "virus infection".

Remark 2: The philosophical discussion of dichotomies concerning kinds of knowledge, i.e. "terminological versus empirical" or "analytical versus synthetical" is not covered in this paper. Of course, concept systems need the inclusion of many concepts related partitively to other concepts (e.g. hand is part of arm), what is empirical knowledge. Of course, concept systems need further empirical knowledge in concept definitions in order to be useful. However, when speaking about formalized concept systems, the definitions should focus on terminological (intrinsic) knowledge as much as possible. Otherwise, sound and complete and tractable algorithms for the processing of formalized concept systems are not ensured.

The most important message of this remark is that concept systems are aiming at knowledge organization. This property of concept systems is what research groups in the fields of medical knowledge processing and medical linguistics are interested in.

Term (ISO 1087 [41]):

Designation of a defined concept in a special language by a linguistic expression.

Nomenclature (ISO 1087 [41]):

System of terms which is elaborated according to pre-established naming rules.

Systematized Nomenclature (as used in this paper):

Nomenclature with a concept system as its basic organization principle.

Remark: Regarding SNOMED, Rothwell [25] writes: "... terms are placed into taxonomic hierarchies expressing their natural relationships to one another." The emphasis on a group of interrelated terms is different from the emphasis on concept systems. Concepts are assumed to be non-linguistic abstract entities that are expressed by linguistic terms. We acknowledge the significance of this difference. However, in this paper we focus mainly on the ordering principles underlying nomenclatures; hence, on concept systems.

Indexing (as used in this paper):

Semantic characterization of language data by one or more concepts of a nomenclature. Usually, a translation of the language data into a formal expression based on the concepts is needed, in order to achieve even a meaning-preserving reconstruction of the language data.

Standardization (as used in this paper):

Commitment to a representation of language data based on systematic nomenclatures or concept systems inclusive a representation language. Both, the nomenclature and the representation language should be as widely accepted as possible.

Remark: Standardization can be carried out by indexing of already entered medical language data on the one hand, and by structured data entry on the other hand. A wide acceptance can be credited to nomenclatures like SNOMED, however, for representation languages such an acceptance is missing. Of course, this definition of standardization is very restricted.

Example: SNOMED - Systematized Nomenclature of Human and Veterinary Medicine [35] - is a multi-axial systematized nomenclature, where medical statements can be built by combinations of concepts from the different axes.

Remark: In this paper we treat SNOMED as a concept system augmented with valuable term-related information. SNOMED offers a meta-language for the meaningful representation of complex concepts including general linkage modifiers. It incorporates also a disease axis or module that seems to be a classification. The entries in this axis have links to the other modules and to ICD-9-CM-classes. We have some concerns with the disease axis as a classification, at least with respect to the definition below.

Quote from Coté, et al. [35]: "SNOMED, the Systematized Nomenclature of Medicine, is a structured nomenclature and classification of the terminology used in human and veterinary medicine. ... SNOMED International possesses many of the features needed for knowledge rèpresentation in medicine. It is, in fact, a data structure that is modular, open-ended and possesses a flexibility suitable for expressing features that we believe are necessary for a well-grounded medical terminology that is capable of 'packaging' concepts, i.e., information units, into computer processable entities."

Coding examples:

Tuberculosis of the right main bronchus (DE-14814) = Right main bronchus (T-261) + Granuloma (M-44) + M. tuberculosis (L-21801) + Fever (F-03003)

Gonococcal meningitis (DE-12091) = Meninges, NOS (T-A111) + Inflammation, NOS (M-4) + Neisseria gonorrhea (L-22201) + Sexually transmitted disease (DE-016)

Remark: It is seen from the examples that some concepts are explicitly transferred from the original phrase (e.g. T-261), some are added from the human or computer-supported indexer (e.g. F-03003), and some are added by provided links within SNOMED (e.g. DE-016). standardized data it is easy to retriev diagnoses concerning "Bronchial disease even "Diseases of the right lower lobe m segmental bronchus (T-26420)" when ne Standardized data can be easily aggreg Sound statistical analyses, however, ne classification of data. For classifying base standardized data all class-relevant diagi have to be defined explicitly (e.g. for I(089.8 "Gonococcal infection of other spec sites", see below). The availability of (relevant criteria like "isolated" in ICD9 (has to be ensured. The use, either of expl given data or of inferred data for the assignı to a class, makes a difference statistically.

Class (ISO 1087 [41]), interpreted as concept's extension:

Totality of all objects to which a con refers.

Rubric (as used in this paper):

Medical phrase for designating and iden ing a class.

Classification (ISO 5127, Part 6 [42]):

Arrangement of concepts into classes their subdivisions to express semantic relati between them; the classes are represented means of a notation.

Partitive Classification:

A classification that establishes an exhau ve and exclusive partition of a given domain.

Remark: From a general point of vi ideally given objects can be classified uniqu to one class, e.g. all living people according their age without saying something about ı acquisition of this information.

(Statistical) Classification (as used in t paper):

A partitive classification that meets ı following conditions:

a) The fundamental divisions must be cle and unambiguous at every stage.

b) The divisions must be exhaustive at eve stage.

c) The divisions must be exclusive at eve stage.

d) Whenever possible, the compartments each division should be of comparable rank.

e) At each division there should be a sing and uniform fundamental division.

f) The divisions should be justifiable at eac step in terms of one single governing purpo throughout the classification as a whole.

g) Classifications should be informati (have a clear purpose).

Remark 1: This definition is oriented ı Rescher [43, S.52-54]. Analogous to the classif cation of quantitative data according to the statistical distribution, qualitative data shoul be classified with respect to the expecte frequencies of the value categories. Classific tions must be reflected within the field c statistical experiments. For such experimen methodological requirements have to be re garded, e.g. the definition of the populatio under investigation, precise formulation of th question(s) of interest, and an unbiased sampl ing. We agree almost with Côté who said i [24]: "That infers also that the ICD, includin the ninth revision, is not a 'true' classification o diseases, and that the term 'statistical classifica tion' should be defined as a list of statisticall significant entities and groups of entities base

). With
eve all
ases" or
medial
needed.
regated.
need a
based on
iagnoses
ur ICD9-
specified
of class-
ICD9 011.3
of explicitly
assignment
ally.
ted as the

a concept

nd identify-

42]):
lasses and
relations
esented by

in exhausti-
domain.
nt of view
ed uniquely
according to
about the

used in this

meets the

ust be clear

e at every

e at every

rtments of
e rank.
be a single

able at each
ng purpose
ole.
informative

riented to
the classifi-
ng to their
data should
e expected
. Classifica-
ne field of
xperiments
to be re-
population
ation of the
sed sampl-
ho said in
including
fication of
classifica-
tistically
es based

predominantly on prevalence or considered importance." With 'true' classification he seems to designate concept systems.

Remark 2: Opposed to pure partitive classifications, statistical classifications are not dealing with ideally given objects. Objects are classified indirectly through what is observable and what is interpretable from already recorded observations. Hence, classifications incorporate classes like "unspecified individuals with respect to division criteria" and other meta-instructions referring to the act of observing and recording.

Remark 3: Opposed to concept systems or systematic nomenclatures that are provided primarily for a patient specific standardization, classifications are primarily population oriented. However, there are exceptions like the TNM classification [14], where besides the support of sound clinical cancer research, there is also a patient-specific aspect. Dependent on a precise staging and grading of tumors according to the classification rules, a prognosis can be derived with consequences for the ongoing diagnostic and treatment of a patient.

Classifying (as used in this paper):

Unique assignment of an object (denoted by a medical phrase) to one class of a classification.

Example: ICD - Int. Classification of Diseases, Injuries and Causes of Death [44] is a mono-axial, statistical classification.

Quotation from WHO [44], see also [45]:

"A statistical classification of disease must be confined to a limited number of categories which will encompass the entire range of morbid conditions. The categories should be chosen so that they will facilitate the statistical study of disease phenomena. A specific disease entity should have a separate title in the classification only when its separation is warranted because the frequency of its occurrence, or its importance as a morbid condition, justifies its isolation as a separate category. On the other hand, many titles in the classification will refer to groups of separate but usually related morbid conditions. Every disease or morbid condition, however, must have a definite and appropriate place as an inclusion in one of the categories of a statistical classification."

Coding examples:

Tuberculosis of the right main bronchus → 011.3 Bronchial tuberculosis excl. Isolated bronchial tuberculosis (012.2)

Gonococcal meningitis → 098.8 (†) Gonococcal infection of other specified sites (classifying primarily the cause aspect) → 320.7 (*) Meningitis in other bacterial diseases classified elsewhere (manifestation aspect)

Remark: It is seen from the examples that classifying generally means loss of information. With classified data it is not easy to retrieve all diagnoses concerning "Bronchial diseases" at all. They are classified into classes like the two given in the example, but also into 192.1 "Malignant neoplasms of meninges", 741 "Spina bifida" etc. All these classes include diseases of the bronchus, not specified in more detail. Sound statistical analyses can be easily carried out inherently. However, given a goal of evaluation, conflicting with the classification's goal, sound statistical evaluations are no longer possible. For example, disease-oriented cost calculations cannot be carried out based on the mortality-oriented ICD. A class like 410 "acute myocardial infarction" is inhomo
according to costs.

(Controlled) Medical Vocabulary (IS
[41] and CEN ENV 12264 [46]):

Terminological dictionary containi
restricted to) the terminology of a spec
ject field or of related subject fields an
on terminological work.

Remark: The term "controlled med
cabulary" does not refer to any st
properties or ordering principles underl
vocabulary. It is used more as a place to
all kinds of terminological systems lik
saries, nomenclatures, thesauri, and cl
tions. Not the schemas, but the str
collection of lexical units, terms or con
the instances – are in the focus of the de

Coding System (CEN ENV 12264 [46

A combination of a system of con
terminology (rubrics), a set of code valu
at least one coding scheme to relate the c
the concepts, terms or classes.

Remark: Most of the existing m
created vocabularies use codes for ide
terms, concepts or classes uniquely. F
more, hierarchical coding schemes
hierarchical relations between conce
classes by string relations (e.g. M-885 is
with all the restrictions explained in this
It is essentially the absence of this prope
that codes are used for predefining
relations as in existing nomenclature
makes them different to concept systems.
forming medical language data into
systems is uniformly called "coding", reg
of further refinements like indexin
classifying.

Exhibit I

Compositional and Enumerative Designs for Medical Language Representation

Anne-Marie Rassinoux[1], Ph.D., Randolph A. Miller[1], M.D.
Robert H. Baud[2], Ph.D., Jean-Raoul Scherrer[2], M.D.
[1]Division of Biomedical Informatics, Vanderbilt University, Nashville, TN
[2]Medical Informatics Division, University Hospital of Geneva, Switzerland

Medical language is in essence highly compositional, allowing complex information to be expressed from more elementary pieces. Embedding the expressive power of medical language into formal systems of representation is recognized in the medical informatics community as a key step towards sharing such information among medical record, decision support, and information retrieval systems. Accordingly, such representation requires managing both the expressiveness of the formalism and its computational tractability, while coping with the level of detail expected by clinical applications. These desiderata can be supported by enumerative as well as compositional approaches, as argued in this paper.

These principles have been applied in recasting a frame-based system for general medical findings developed during the 1980s. The new system captures the precise meaning of a subset of over 1500 medical terms for general internal medicine identified from the Quick Medical Reference (QMR) lexicon. In order to evaluate the adequacy of this formal structure in reflecting the deep meaning of the QMR findings, a validation process was implemented. It consists of automatically rebuilding the semantic representation of the QMR findings by analyzing them through the RECIT natural language analyzer, whose semantic components have been adjusted to this frame-based model for the understanding task.

INTRODUCTION

Medicine is a domain involving a huge amount of information, most of which is still expressed through textual forms. Understanding and extracting the meaning embedded in these texts is a continuous challenge to researchers in medical informatics[1]. Standardization efforts towards reducing the expressiveness and peculiarities inherent in medical language have led to the emergence of two major methods of organizing medical information. On the one hand, different thesauri or controlled medical vocabularies (CMVs) - such as the UMLS Metathesaurus or the ICD classification - are now available, affording an extensive set of relevant terms to express patient-specific observations. On the other hand, more formal semantic models for medical concept representation - such as the Medical Entities Dictionary (MED) or the GALEN model - have come to light, fostered by the need to transcend words and phrases and to capture their "meaning". These conceptualization efforts result in language-independent and compositional systems for modeling the intricate concepts of medicine.

The counterbalancing features underlying the two approaches for medical concept representation relate mainly to breath of coverage and depth of representation, which respectively involve enumerative and compositional strategies. Actually, CMVs allow rapid and easy incorporation of new terms without disturbing the general representational architecture. But, enumerative description performed through language-surface form entails redundancy and inconsistency which can impede the overall maintenance of such vocabularies. Besides, representing medical concepts in a more computationally meaningful manner implies decomposing and structuring information in a formal way, which is suitable for manipulation by computer programs. This constitutes a more labor-intensive and time-consuming task. Therefore, it is necessary to limit the medical subject domain for fine modeling to yield concrete outcomes in a reasonable period of time.

This paper presents a challenging effort undertaken by the authors to recast the frame-based system initially developed by Miller, Masarie, et al.[2,3]. The objective is to obtain a more computationally tractable model which introduces conceptual graphs[4] to represent and standardize the various compositional aspects of medical information. The checking and adjustment have been manually performed for 750 generic medical finding frames that capture the meaning of 1500 selected QMR "surface-level" findings. One way to validate the accuracy and tractability of the new frame-based system is to use Natural Language Processing (NLP) techniques to check the meaning of QMR findings against this system.

BACKGROUND TO THE FRAME-BASED SYSTEM

Characteristics of the QMR Vocabulary

The QMR vocabulary (which is a superset of the original INTERNIST-I vocabulary)[5] was created to describe possible (reported) patient findings in diseases in general internal medicine. It contains over 4500 clinical manifestations, including patient symptoms, physical findings, and laboratory test results. This vocabulary was derived from extensive manual literature review and serves the purpose of providing input for the QMR diagnostic program[5]. Such a vocabulary fits the characteristics of enumerative systems. Terms are mainly described through noun phrases consisting principally of medical phrases with generally accepted definition and usage, as shown in Figure 1.

Finding: DYSPNEA PAROXYSMAL NOCTURNAL
This phrase corresponds to the medical expression "paroxysmal nocturnal dyspnea", which describes an *acute onset of inappropriate shortness of breath or similar difficulty in breathing occurring at night.*

Finding: ORTHOPNEA
This term describes a *discomfort in breathing which is brought on or aggravated by lying flat.*

Figure 1 - QMR findings and their clinical definition

Moreover, it is worth noting that the language used to express these findings is strongly stereotyped and has not strictly applied the syntactic formative rules of English. In particular, conventional orders of certain words are not followed (in order to maintain a new form of internal consistency for word order), and prepositions are less frequently used. These surface observations already suggest that semantic categories appear to be more appropriate to determine the details of interpretation of these noun phrases as syntax is used in a "fancy" way.

Evolution of the Frame-Based System

In order to capture the clinical complexity of the QMR findings, Miller, Masarie et al. developed a frame-based interlingua[2, 3], which has been further used to facilitate the translation between CMVs. This system limits itself to collecting - through a bottom-up approach reviewing each existing QMR finding - a core set of central concepts considered as relevant to recognize any and all sensible information embedded in the QMR findings. For this, it is assumed that any clinically relevant statement about patients contains at least one identifiable central concept. Figure 2 shows an example of a generic frame, followed by the list of QMR terms which are candidate to map this structure.

DYSPNEA
Generic Frame: last edited on */*/* by *****
Allowable Status: Presence Or Absence
Normal Status: Absent
Method(s)
Name: Cardiopulmonary Observation
Reliability: 4
Qualifier(s)
Pattern Of Occurrence; Time Duration Qualitative; Time Duration Quantitative; Influence On Dyspnea; Time Of Day; Time Onset Qualitative

DYSPNEA ABRUPT ONSET
DYSPNEA ACUTE RECURRENT ATTACK <S> HX
DYSPNEA AT REST
DYSPNEA AT REST RELIEVED BY RECUMBENCY
DYSPNEA EXERTIONAL
DYSPNEA IMPROVEMENT AFTER HEMOPTYSIS HX
DYSPNEA PAROXYSMAL NOCTURNAL
ORTHOPNEA
DYSPNEA RELIEVED BY SQUATTING HX

Figure 2 - Initial generic frame structure

The generic frame structure provides the backbone for describing the fundamental characteristics associated with the central concepts. This structure integrates both the status description of the considered medical concept (i.e. its "default normal value", usually describing clinical findings as normal or abnormal conditions affecting anatomical sites) and the methods used to elicit such a concept in a medically meaningful fashion, as well as the potential qualifiers which can be applied to this central concept. The qualifiers lists (also called item lists[3]) are useful to encapsulate fine details. Such qualifiers are maintained apart from the generic frames as they specify well-defined features often applicable across a number of generic frames. The qualifiers description incorporates both a limited set of values as well as a header stating the logical relationship among the components. For example, the qualifier *'Time Duration Qualitative'* is represented through the following values: *Acute, Subacute, Chronic* linked by the header *ExactlyOneOf.*

The thorough and enumerative method used to build the frame-based system insures the richness and accuracy of the resulting model. Indeed, the builder of the knowledge base system (usually referred to as the expert) was concerned primarily with the extraction of relevant concepts from the test set of QMR terms (and some terms from DXplain and HELP as part of the UMLS project) without being compelled to apply some protocol instructions. However, this approach limited development of a fully language-independent and computationally tractable system of medical concept representation. On the one hand, it appears that the concept system itself is not clearly separated from the precise language used for specifying its components. The extensive use of complex linguistic names to label central medical concepts (such as '*Left Ventricular End Diastolic Internal Diameter*'), as well as qualifiers (such as *'Timing Within Systole Or Diastole'*) blurs the separation between the concepts to be represented, and the linguistic terms and mechanisms used to refer to those concepts. Moreover, the separation between concepts and relationships is masked by the use of equivocal labels (such as *'Influence On Dyspnea'*). On the other hand, the flat enumeration of generic frames, making use of a large amount of conceptual entities which are not structured in a hierarchical framework, causes trouble for maintenance and navigation through the system itself.

Facing these drawbacks, a new structure[6] has been developed by the authors. The result, based on the example shown in Figure 2, is displayed below in Figure 3.

```
genericFrame('Dyspnea',
  [existentialStatus:
          [allowableStatus('PresenceOrAbsence'),
          normalStatus:absent],
   definition: ['Difficulty', [actsOn('Breathing')]],
   methods:['CardiopulmonaryObservation'('4')],
   qualifiers:[hasProcessPattern('ProcessPattern'),
             hasChronicity('Chronicity'),
             hasDuration('Duration'),
             isInfluencedBy(['Hemoptysis','Exercise',
                             'BodyPosition','Rest']),
             occursDuring('TimeOfDay'),
             hasOnset('TypeOfOnset') ] ] ).
```

Figure 3 - Revised generic frame structure

For the sake of clarity, the nature of the manipulated information is highlighted by considering two kinds of generic frames, differing by the type of their status. On the one hand, the *existential frames* describe findings which may or may not occur for a given patient. On the other hand, the *quantitative frames* describe clinical parameters which can be measured. Moreover, except for the slot *qualifier*, the other slots contain mandatory information that helps in recognizing, in a non-ambiguous way, the current generic frame.

DEALING WITH COMPOSITIONALITY

As emphasized in Figure 3, the recasting of the frame-based system dealt mainly with transforming a rather enumerative description into a more structured system, which fits most of the desiderata highlighted by Cimino[7]. The main innovations and their issues are discussed below.

Hierarchy of Concepts

Even if the frame structure used to represent the central medical concepts is convenient to express a first level of description (through slots and fillers), allowing then the initial structure to be inverted according to some criteria, this representation is nevertheless not easy to maintain. Therefore, a hierarchically-structured view of, at least, all the primitive concepts which are useful to describe more complex medical information has been implemented. The high level of this multiple hierarchy (i.e. lattice) first delimits conceptual entities from relationships, thus determining straight-away the atomic objects handled by any compositional process. Second, it separates medical concepts from the modifiers which serve to precisely describe these concepts. Such a subclassification reflects the two main parts of the frame-based system (i.e. the generic frame structure and the qualifiers description) and allows for specifying the weight given to the information, in particular for its potential use by NLP tools. In addition, the part of the hierarchy listing the methods is especially detailed, as such methods play an important clinical role in eliciting the central concepts.

Formal Definitions

In order to be able to exploit (with a computer) the meaning of complex medical expressions, formal definitions are introduced. At this level, it is important to delineate definitional knowledge from assertional knowledge[7]. The literal definition, added in the frame-based system, only reflects the terminological (also called lexical or literal) meaning embedded in the central concept name. For example, the concept *Dyspnea* refers to *a difficulty* (Greek prefix "dys") *in breathing* (Greek root "pnea"). Such a definition, acting as definitional knowledge, is often not complete enough to describe the full clinical meaning of the treated concept. This meaning, referring to the assertional (also called encyclopedic or contextual) knowledge is explicitly expressed in the model itself (through the slots methods, qualifiers...), which establishes the context and circumstances in which the central concept should occur in the clinical reality.

This literal definition presents some interesting features. First, it is expressed through the Conceptual Graph (CG) formalism[4], which allows a convenient graphical representation of concepts linked through relationships. This formalism offers a rich representation as conceptual graphs can be arbitrarily large. It also supports various kinds of operations, in particular, contraction and expansion, which are especially important in handling definitions. Second, this definition is of paramount importance in retrieving the different linguistic expressions of the central concept from textual documents, especially when this concept is expressed with a multi-word phrases (that is to say, consisting of more than one word). For example, the definition related to the central concept *Dyspnea* (see Figure 3) is heavily relevant to extract this concept from the sentence *"The patient presents some difficulties in breathing at night."*, using a semantic-oriented medical language processor such as the RECIT system[8].

Finally, having a compositional model allows equivalent definitions to be expressed and maintained at the conceptual level, thereby eliminating the need to provide the system with numerous lexical variants, as discussed in the next section.

Hierarchy Annotation

The hierarchy annotation is particularly important for ensuring that tools with access to textual sources, such as retrieval engines or natural language processors, function correctly. Indeed, it consists in an extensive enumeration of synonyms and related terms (expressed through single words or multi-word phrases) which are used to refer to concepts, and are stored in the so-called dictionaries. For example, the concept *Dyspnea* can be annotated in English by the following linguistic expressions: *dyspnea*, *breathlessness*, *shortness of breath*, etc. However, noun phrases such as *discomfort in breathing* are not mandatory, as long as the literal definition of *Dyspnea*, as well as the annotations of the primitive concepts composing this definition, are provided by the system.

Scope of Relationships

An important aspect of compositionality is handled through the notion of relationships and is emphasized in the frame-based system by replacing qualifiers such as '*Influence on Dyspnea*' into the relationship *isInfluencedBy* which points to the set of relevant concepts (see Figure 3). At this level, it is important to clarify and enforce the scope of relationships in order to avoid misinterpretation of the meaning specified in the generic frames. For instance, the generic frame *AbdominalPain* (whose literal definition is *[Pain,[hasLocation(Abdomen)]]*) embeds the qualifier *Periodicity,* whose *Colicky* is a possible value. Saying that an *AbdominalPain* can be *Colicky* does not infer any information on the concept *Pain* in particular. Therefore, the relationship *hasPeriodicity* must strictly link the full concept *AbdominalPain* (and not part of its definition) to the possible value *Colicky* in order to avoid wrong interpretation.

EVALUATION OF THE FRAME SYSTEM

The evaluation of the frames' content is essential for the consistency, extension, and sharing of the overall system. The global process is reported in Figure 4, and is discussed according to the expressiveness and computational tractability of the revised frame system.

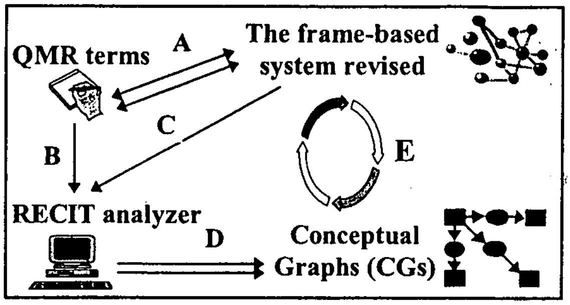


Figure 4 - The global evaluation process

The way the frames were created[3], and then reviewed[6] - by checking each frame's content to the set of QMR terms, candidates to be instantiated through this frame (link A in Figure 4) - constitutes a first validation of the expressiveness of the system in grasping the QMR terms meaning. This manual validation is then reinforced through the use of the RECIT multilingual analyzer[8], which automatically analyzes and stores the semantic content of the QMR terms under the form of CGs (link D in Figure 4). This last process, useful to validate both the granularity and tractability of the frame system is presented below.

The RECIT analyzer first applies "Proximity Processing" rules to group neighboring words together, and second links these semantic fragments into a sound structure expressed through conceptual graphs. For the task of analyzing the QMR terms (link B in Figure 4), the semantic components of the RECIT analyzer have been grounded directly from the revised frame-based system (link C in Figure 4). This latter connection emphasizes the computational tractability of the frame system, and has been facilitated by our previous experience in relying on the GALEN model[8]. For the present experiment, the generic frames are integrally used as valid conceptual schemata, useful to accurately build the sound representation of medical sentences. The compatibility rules used in the first analysis phase are also extracted from this structure.

RESULTS

As a preliminary test, 200 QMR findings, instantiating nearly 50 generic frames, were given as input to the RECIT system. The results were reviewed for the two analysis phases. In particular, failures during the proximity processing phase generally occur because of lack of specifying a particular concept (as the concept *Orthopnea*, further classified in the hierarchy as a child of the concept *Dyspnea*), or lack of a specific annotation for an already existing concept (as the word *nocturnal* which annotates the concept *Night* defined as a particular value of the qualifier *TimeOfDay*). Moreover, failures in producing a unique conceptual graph in the second phase of the analysis process, which points to a generic frame, clearly reflect the need to add a literal definition (as for the concept *Orthopnea* relating to *breathlessness lying flat*), or to specify a new relationship in the generic frame structure (see Figure 3). Such a refinement process of the hierarchy, dictionaries, literal definitions, and generic frame structure, can be considered as a feedback loop from the NLP system to the model as illustrated by link E in Figure 4.

Finally, as the model evaluation lies on the result of the RECIT analyzer, the performance of this analyzer toward dealing with the medical jargon has also been readjusted, facilitated by the fact that such analyzer was specifically designed for this task[8].

CONCLUSION

This paper reassigns the importance of compositional and enumerative designs for medical language representation, respectively between the modeling process and the linguistic annotation process (which underlies any concept model intended to be used by some NLP tools[8]). It clearly emphasizes the benefits of managing a fully compositional and tractable model of medical concept representation, in parallel with an enumerative dictionary of synonyms and related terms, in order to handle the intricacy of the medical language.

The automatic validation process of the frame-based system, using the RECIT medical language analyzer, allows both the expressiveness and tractability of the model to be checked. This experiment promotes NLP tools, whose generation has also been successfully applied for this task[9], as quality assessment processes of concept models.

Acknowledgments

This work is supported by grant number 8220-046502 from the "Fonds National Suisse de la Recherche Scientifique". Work on the generic frame schema was originally supported through NLM contract N01-LM-6-3522. Dr. Miller's current work is supported in part by NLM contract 1 R01 LM06226-01A1.

References

1. Spyns P. Natural Language Processing in Medicine: An Overview. Meth Inform Med, 1996; 35(4/5): 285-301.
2. Miller RA. A Computer-based Patient Case Simulator. Clin Research, 1984; 32: 651A.
3. Masarie FE, Miller RA, Bouhaddou O, Giuse NB, Warner HR. An Interlingua for Electronic Interchange of Medical Information: Using Frames to Map between Clinical Vocabularies. Comput Biomed Res, 1991; 24(4): 379-400.
4. Sowa JF. Conceptual Structures: Information Processing in Mind and Machine. Reading, MA: Addison-Wesley Publishing Company, 1984.
5. Miller RA, Masarie FE, Jr. Use of the Quick Medical Reference (QMR) Program as a Tool for Medical Education. Meth Inform Med, 1989; 28(4): 340-345.
6. Rassinoux A-M, Miller RA, Baud RH, Scherrer J-R. Modeling Principles for QMR Medical Findings. In: Cimino JJ (ed). Proceedings of the 1996 AMIA Annual Fall Symposium (formerly SCAMC). Philadelphia: Hanley & Belfus, Inc. 1996: 264-268.
7. Cimino JJ. Desiderata for Controlled Medical Vocabularies in the Twenty-First Century. Proceedings of the Fourth International Conference on Natural Language and Medical Concept Representation, Jacksonville, Florida, January 19-22, 1997: 257-267.
8. Rassinoux A-M, Wagner JC, Lovis C, et al. Analysis of Medical Texts Based on a Sound Medical Model. In: Gardner RM (ed). Proceedings of SCAMC 95. Philadelphia: Hanley&Belfus, Inc., 1995: 27-31.
9. Baud RH, Rodrigues J-M, Wagner J, et al. Validation of Concept Representation Using Natural Language Generation. Proceedings of the 1997 AMIA Annual Fall Symposium (formerly SCAMC). Nashville, TN, October 25-29, 1997.